

# Emergens og autonomi

Det mentales status i minimal fysikalisme

Hovedfagsoppgave i filosofi  
Filosofisk institutt  
Universitetet i Oslo

Av Øistein Schmidt Galaaen  
veileder Bjørn Ramberg  
Høsten 2001

# Innholdsfortegnelse

<b>Forord</b> .....	s.1
<b>0 Innledning</b>	
0.1 Problemstillingen .....	s.2
0.2 Analytisk innholdsfortegnelse .....	s.10
<b>Del I: Problemet for psykologiens status</b>	
<b>1 Psykologi og fysikk er vesensforskjellige disipliner</b>	
1.1 Innledning .....	s.12
1.2 Karakterisering av psykologi og fysikk .....	s.12
1.3 Forskjellen mellom psykologi og fysikk .....	s.15
1.4 Avsluttende kommentarer .....	s.20
<b>2 Minimal fysikalisme</b>	
2.1 Innledning .....	s.21
2.2 Kravene til minimal fysikalisme. Svak og global superveniens .....	s.22
2.3 Sterk superveniens .....	s.25
2.4 Mereologisk superveniens .....	s.28
2.5 ”Appendiks”: Forholdet mellom superveniens og avhengighet .....	s.31
2.6 Oppsummering .....	s.33
<b>3 Intuisjon om realisme</b>	
3.1 Innledning .....	s.33
3.2 Egenskaper og kausale krefter .....	s.34
3.3 Prinsippet om det mentales realitet .....	s.40
<b>4 Problemet for psykologiens status</b>	
4.1 Innledning .....	s.41
4.2 Konflikten mellom fiksering og autonomi .....	s.41
4.3 Emergentisme kan tilsynelatende løse konflikten .....	s.43
<b>Del II: Emergentisme</b>	
<b>5 Emergensbegrepet</b>	
5.1 Innledning .....	s.47
5.2 Den epistemologiske komponenten .....	s.48
5.3 Den ontologiske komponenten .....	s.53
5.4 ”Downward causation” er fundamental .....	s.56
5.5 Formulering av emergensbegrepet .....	s.59

### **Del III: Kims kritikk av emergentisme**

#### **6 Utgangspunktet for et forsvar mot Kims kritikk av emergentisme**

6.1 Innledning.....	s.60
6.2 Innledende om problemer for emergentisme .....	s.60
6.3 Generelt om Kims kritikk av emergentisme.....	s.62
6.4 Kims kritikk av den epistemologiske komponenten .....	s.63

#### **7 Kims argumenter mot ”downward causation”**

7.1 Innledning .....	s.67
7.2 Kims superveniensargument mot ”downward causation” .....	s.67
7.3 Superveniensargumentets forutsetninger .....	s.71
7.4 Prinsippet om det fysiske domenes kausale lukkethet .....	s.75
7.5 Prinsippets implikasjoner .....	s.80
7.6 Konsekvenser av analysen av superveniensargumentet .....	s.81
7.7 Kims asymmetribaserte argument mot ”downward causation” .....	s.81

#### **8 En vei utenom Kims argumenter mot ”downward causation?”**

8.1 Innledning .....	s.84
8.2 Sperrys hjul I: ”Downward causation” uten interferens med mikrostrukturen .....	s.85
8.3 Sperrys hjul II: Arbeidsdeling mellom mentale og fysiske egenskaper .....	s.88
8.4 Problemer med arbeidsdelingsmodellen. Van Gulicks mønstre .....	s.89
8.5 Platonisk reifisering? Intuisjoner og vitenskapelig praksis .....	s.93

<b>9 Epilog .....</b>	<b>s.95</b>
-----------------------	-------------

<b>10 Bibliografi .....</b>	<b>s.97</b>
-----------------------------	-------------

## **Forord**

Denne oppgaven er resultatet av to semestres arbeid ved filosofisk institutt ved Universitetet i Oslo, våren og høsten 2001. Jeg vil rette en spesiell takk til Bjørn Ramberg for tålmodig veiledning og stimulerende diskusjoner. Jeg vil også takke de andre ansatte og studenter ved filosofisk institutt, som har hjulpet meg gjennom seminarer, diskusjoner eller på andre måter. Ingen nevnt, ingen glemt.

## 0 Innledning

Denne oppgaven handler om følgende generelle problem: Hvordan kan psykologi utgjøre et autonomt begrepslig perspektiv på verden, når fysikalistiske forpliktelser samtidig tilsier at denne verden er grunnleggende fysisk? Jeg vil motivere dette problemet ved å vise hvordan det oppstår fra følgende forutsetninger: (1) Psykologi og fysikk er vesensforskjellige begrepslige praksiser. (2) Ontologiske forpliktelser (såkalt minimal fysikalisme) tilsier at verden er grunnleggende fysisk: Alle objekter har fysiske egenskaper og objektenes mentale egenskaper er avhengige av, og bestemt av disse fysiske egenskapene. (3) En form for egenskapsrealisme som tilsier at psykologien må referere til kausalt virksomme mentale egenskaper i verden. Deretter vil jeg forsøke å rimeliggjøre at såkalt emergentisme kan utgjøre en løsning på dette problemet. Emergentisme tilsier kort sagt at det finnes emergente egenskaper ved visse komplekse objekter. Disse egenskapene er egenskaper som ikke innehaes av objektenes deler, og som i en bestemt forstand ikke kan forutsies ut fra viten om objektenes delnivå. Et sekundært mål i denne oppgaven er – gjennom diskusjonen – å si noe generelt om forholdet mellom metafysiske forutsetninger og intuisjoner i dette området av bevissthetsfilosofien. Jaegwon Kim står sentralt i diskusjonen, og særlig er hans *Mind in a Physical World* en hovedkilde for oppgaven.<sup>1</sup>

### 0.1 Problemstillingen

Jeg vil her gi en kort og intuitiv fremstilling av hvordan problemet oppstår, samt av emergentisme som en mulig løsning. I Del I av oppgaven (kap.1-4) utvikler jeg problemet på en fyldigere måte. Fra dette vil det fremgå at emergentisme er en mulig løsning av problemet. Del II (kap. 5) er viet til emergensbegrepet, og del III (kap. 6-8) til diskusjon av argumenter mot emergentisme. I denne innledningen vil jeg, på en løselig og intuitiv måte, bruke noen begreper, som jeg senere vil gjøre nærmere rede for.

Den første forutsetningen for problemet er: (1) *Psykologi og fysikk er vesensforskjellige begrepslige praksiser*. Mennesket er et vesen med språkevne, og forholder seg til seg selv og den verden det bebor gjennom begrepene det bruker. I menneskets språk skiller to vokabularer seg ut: For mennesket bruker fysiske begreper om alle objekter i verden – også om seg selv. Men om seg selv bruker det også *mentale* begreper. Mennesket har altså

---

<sup>1</sup> Kim (1998a). Heretter refererer jeg til denne boken på en spesiell måte: (MPW 12) viser f.eks. til s.12 i den. I MPW formuleres hovedtrekkene i Kims reduksjonistiske posisjon. Det vil imidlertid ofte bli nødvendig å gå til artikler der han argumenterer mer utførlig og eksplisitt.

en forståelse av seg selv både som et *psykologisk* og som et *fysisk vesen*. Det er imidlertid en vesensforskjell mellom de begrepslige praksisene som er knyttet til disse forståelsesmåtene.

Denne forskjellen formuleres godt i Donald Davidsons artikkel "Mental Events" (1970): Hans utgangspunkt her er at hendelser er mentale forsåvidt som de kan gies en mental beskrivelse, dvs. en beskrivelse som inneholder et mentalt verb som "tenker" eller "frykter." (Tilsvarende er hendelser fysiske forsåvidt som de kan gies en fysisk beskrivelse.) Det vil bemerkes at slike mentale verb er intensjonale verb. Kjennetegnet ved en psykologi som benytter seg av slike mentale beskrivelser blir dermed intensjonalitet.<sup>2</sup> At psykologi slik karakterisert er intensjonal, betyr at den benytter seg av såkalte proposisjonale attityder, eller påstandsholdninger. Disse uttrykker måter personer kan forholde seg til en påstand p på, som f.eks. "Ole *håpet* at p."

Davidson argumenterer for at det mentale vokabularet er preget av en holisme. Det utgjør et hele som er avgrenset fra det fysiske vokabularet. Dersom man forsøker å gi mentale begreper definisjoner som bare bruker begreper fra et fysisk eller behavioristisk vokabular, vil man oppdage dette: La oss si at jeg forsøker å definere "Ole tror at det er eksamen om en uke" ut fra beskrivelser som "Ole er disponert for å kaldsvette, lese mye, møte opp på et bestemt sted om en uke etc." Problemet er at dette forutsetter at Ole *ønsker* å ta eksamen (eller i alle fall har en pro-holdning til dette), at han *vet* at eksamen er på det og det stedet, at han *tror* at det er mulig for ham å ta eksamen osv.<sup>3</sup> Disse mentale forutsetningene har i sin tur nye mentale forutsetninger osv., osv. Vår definisjon vil derfor aldri bli fri for mentale beskrivelser.

Denne holismen som avgrenser den mentale begrepslige praksisen fra den fysiske er ifølge Davidson av en spesiell art: For begrepene i de to praksisene har nemlig *vesensforskjellig anvendelsesbetingelser*. Mulighetsbetingelsen for anvendelsen av psykologiske termer er at vi forutsetter en bakgrunn av rasjonalitetsnormer. I det øyeblikk vi f.eks. tilskriver en person en oppfatning, har vi implisitt tilskrevet ham en rekke andre oppfatninger, og vi har forpliktet oss til å forstå ham som stort sett rasjonell og stort sett innehavende sanne oppfatninger. (Dette er Davidsons berømte "Principle of Charity.")<sup>4</sup> Denne betingelsen er rimelig gitt at psykologi er knyttet til vår interesse av å forstå oss selv og andre som *tenkende* og *handlende*, dvs. som rasjonelle aktører.

---

<sup>2</sup> Davidson (1970) s.210-211.

<sup>3</sup> Se ibid. s.217 for et annet eksempel.

<sup>4</sup> Davidson (1970) s.221-22.

På den andre siden forutsetter anvendelsen av *fysiske* termer, slik Davidson ser det, strenge, generelle og lovmessige sammenhenger. Fysiske generaliseringer er av en slik art at man forventer at de (i hvert fall prinsipielt) skal kunne skjerpes til unntaksløse lover.<sup>5</sup> Poenget nå er at psykologi og fysikk er vesensforskjellige hva gjelder de anvendelsesbetingelser de forutsetter og de erkjennelsesinteresser de svarer til. Derfor vil man forvente at psykologi i stor grad er autonom i forhold til fysikk, og at psykologiske og fysiske begreper ikke kan bringes i pene reduktive sammenhenger:

There are no strict psychophysical laws because of the disparate commitments of the mental and physical schemes. It is a feature of physical reality that physical change can be explained by laws that connect it with other changes and conditions physically described. It is a feature of the mental that the attribution of mental phenomena must be responsible to the background of reasons, beliefs and intentions of the individual. There cannot be tight connections between the realms if each is to retain allegiance to its proper source of evidence.<sup>6</sup>

Den metafysiske gulroten som Davidsons posisjon (såkalt anomal monisme)<sup>7</sup> tilbyr, er at den kombinerer denne autonomien (spesielt fravær av strenge lovmessige sammenhenger mellom det mentale og det fysiske) med metafysisk monisme: Enhver mental hendelse er identisk med en fysisk hendelse, fordi enhver mental hendelse også har en fysisk beskrivelse.<sup>8</sup> Men som vi skal se, har vi ontologiske forpliktelser som gjør det mer problematisk om vi på denne måten kan få i både pose og sekk, dvs. få autonomi samtidig som vi tar hensyn til våre ontologiske forpliktelser.

(2) *Minimal fysikalisme*. Kim omtaler minimal fysikalisme som en idé som kan deles av så forskjellige posisjoner som typereduksjonisme og emergentisme. (MPW 15.) I tråd med dette kan minimal fysikalisme karakteriseres som de svakeste metafysiske forpliktelsene vi kan ha, dersom vi vil ta hensyn til det naturvitenskapelige verdensbildet og ikke være substansdualister.<sup>9</sup> Dermed kan minimal fysikalisme være et felles utgangspunkt for diskusjoner om det mentales metafysikk, som kan deles av både reduktive og antireduktive fysikalister. (Når jeg heretter snakker om antireduksjonister vil jeg underforstå at disse er *fysikalister*. Men man kan selvsagt være antireduksjonist uten å være fysikalist.)

Kravene som konkret stilles til minimal fysikalisme er dels *a priori*, og dels empiriske:

---

<sup>5</sup> Ibid. s.219.

<sup>6</sup> Ibid. s.222.

<sup>7</sup> Ibid. s.214

<sup>8</sup> Ibid. s.224.

<sup>9</sup> Vi vil altså unngå ontologier med to typer ting som er vesensforskjellige i den forstand at de ikke deler noen egenskaper. Dersom man vil tenke på en positiv doktrine kan kartesiansk dualisme gjøre nytten: I et berømt tekststed i Sjette Meditasjon skiller Descartes mellom en tenkende (mental) substans ("meg selv som noe tenkende og ikke utstrakt") og en romlig (fysisk) substans ("kroppen som noe utstrakt, ikke-tenkende"). (Descartes (1992) s.73.)

(i) *Monisme*: Velkjente *a priori* argumenter har vist at substansdualisme er en inkonsistent doktrine. Den kan bl.a. ikke gjøre rede for at mentale hendelser forårsaker fysiske hendelser. Videre har vi gode empiriske grunner til å tro at vår substansmonisme skal være en monisme der substansen er den fysiske materie. Dette kan formuleres som at alle objekter har fysiske egenskaper. (ii) *Asymmetri*: Men vi har også empiriske grunner for å si at f.eks. gråsteiner ikke har mentale egenskaper, tvert imot er en stor grad av kompleksitet en betingelse for at et objekt kan ha mentale egenskaper. Så selv om alle objekter har fysiske egenskaper, så er det ikke også slik at alle objekter har mentale egenskaper. (iii) *Avhengighet*: Men empirisk forskning på bl.a. hjernen har lært oss at det er en avhengighetsrelasjon mellom det fysiske og det mentale. Ved å skade eller endre hjernen kan vi fjerne eller frembringe mentale egenskaper. Det er altså ikke vilkårlig at når et objekt har visse fysiske egenskaper så har det også visse mentale egenskaper. Tvert imot er det mentale avhengig av det fysiske i den forstand at det er *nødvendige* sammenhenger mellom fysiske og mentale egenskaper. Dette kan uttrykkes som at når et objekts fysiske egenskaper og verdens fysiske lover er på plass, så er nødvendigvis også objektets mentale egenskaper på plass.<sup>10</sup> (Men fordi det kan være flere fysiske egenskaper som kan spille denne rollen, så gir ikke dette en-til-en sammenhenger mellom mentale og fysiske egenskaper.)

For å gjøre en lang historie kort (jeg skal fortelle den *lange* historien i kap. 2) så tilfredsstiller såkalt sterk superveniens med en underforstått avhengighetsrelasjon krav (i)-(iii), og kan følgelig definere minimal fysikalisme. Ideen om sterk superveniens kan fanges inn på en intuitiv måte ved å si at dersom objekter skal være mentalt forskjellige, så må de nødvendigvis også være fysisk forskjellige:

Mental properties *supervene* on physical properties, in that necessarily any two things (in the same or different possible worlds) indiscernible in all physical properties are indiscernible in mental respects. (MPW 10.)<sup>11</sup>

I tillegg til sterk superveniens forutsetter jeg en ontologi der verden er ordnet lagvis etter del-hele relasjonen i nivåer av stadig økende kompleksitetsgrad: Molekyler er således satt sammen av atomer, celler er satt sammen av molekyler, organismer er satt sammen av celler, samfunn er satt sammen av organismer etc. Egenskaper ved et hele, dvs. makroegenskaper, er selvsagt avhengig av egenskaper ved delene, og måten delene er satt

<sup>10</sup> Her vil det være uenighet om hva man skal inkludere i disse fysiske egenskapene. Enkelte vil ønske å inkludere forskjellige relasjonelle egenskaper mellom objektet og dens omverden, mens andre vil begrense seg til lokale nevrofysiologiske egenskaper. Men hovedpoenget nå er at når disse egenskapene er på plass, så har objektet også nødvendigvis sine mentale egenskaper.

<sup>11</sup> I denne oppgaven er mitt hovedfokus på det intensjonale, og ikke det fenomenale, aspektet ved det mentale. Supervenienstenes må derfor sees på primært som en påstand om intensjonale egenskaper. At såkalte *qualia*, eller fenomenale egenskaper også er superveniente er ikke en absolutt forutsetning i denne oppgaven.



sammen på. (En makroegenskap ved en celle er således avhengig av at cellen består av visse deler – ribosomer og mitokondrier m.m. – satt sammen på en bestemt måte.) Tar du bort delene, eller omorganiserer dem på visse måter, så har du ikke makroegenskapen lenger: Utfør lobotomi på en hjerne, og du vil oppdage at det forandrer hjernens mentale makroegenskaper. Faktisk skal makroegenskapene vise seg å være *sterkt superveniente* på objektets mikrostruktur. (2.4) Dette er også intuitivt rimelig: For setter du sammen to objekter av helt like deler, på en helt lik måte, så vil de nødvendigvis ha de samme makroegenskapene. En annen viktig konsekvens er at det lagdelte verdensbildet gir oss en modell for å snakke om forholdet mellom fysikk og spesialvitenskapene: For mens fysikere studerer objekter av alle kompleksitetsgrader, så har f.eks. kjemikere, biologer og psykologer objekter av bestemte kompleksitetsgrader som sitt studieområde. Vi har altså en ontologi der psykologi har fått sitt eget nivå av virkeligheten: Psykologer studerer mentale egenskaper som innehaes av svært komplekse biologiske objekter, og kanskje bare av biologiske objekter. Så langt virker dette som en oppskrift på autonomi for psykologi: Davidson har lært oss at psykologi fungerer konseptuelt uavhengig av fysikk, og nå har vi også en ontologi der psykologi har fått sitt eget nivå av virkeligheten.

Men sterk superveniens innebærer, som vi skal se mer detaljert senere (2.3), at fysiske egenskaper nødvendigvis er tilstrekkelige for mentale egenskaper. For den eneste måten to objekter kan være mentalt forskjellige på, er ved også å være fysisk forskjellige. Derfor blir det mentale bestemt eller fiksert av det fysiske. Dette skal vise seg å true autonomien.

(3) *Intuisjon om realisme.* I denne oppgaven forutsetter jeg en intuisjon om at psykologi er en vitenskap *om* verden. Dersom psykologiske forklaringer skal ha realitet må derfor dens begreper og prinsipper på en eller annen måte kunne *føres tilbake til* reelle trekk ved verden. Mer konkret vil jeg forutsette at når psykologi forklarer en handling ved å vise til aktørens grunner (spesielt ønsker og oppfatninger) for handlingen, så er dette en kausalforklaring: Grunnen er *årsaken* til handlingen. På dette grunnlaget kan vi se en måte vi kan føre psykologiske beskrivelser tilbake til trekk ved verden: For anta at vi har en psykologisk beskrivelse av handlingens grunn, som referer til en mental egenskap ved aktøren. Anta videre at denne egenskapen faktisk forårsaker handlingen.<sup>12</sup> I så fall har vi en klar tilbakeføring av psykologiens begreper til verden, og slike tilbakeføringer sikrer realitet for psykologiske forklaringer. Det stilles dermed følgende krav til psykologiens realitet: Det

---

<sup>12</sup> Strengt tatt kan egenskaper bare forårsake noe for såvidt som de er instansiert av et objekt. Det er egenskapsinstansieringer og ikke egenskaper som er årsaker. (Jf. MPW 41-42.) Egenskaper alene er ikke noe sted i tid-rom, og følgelig kan de bare sies å forårsake noe alene innenfor et ekstremt platonisk perspektiv. Jeg vil likevel for enkelhets skyld ofte snakke om egenskaper som årsaker og virkninger.

må finnes mentale egenskaper som er kausalt virksomme. Intuitivt innebærer dette ganske enkelt at vårt mentale liv kan påvirke verdens gang.

For å komme til problemet i denne oppgaven så ønsker vi i utgangspunktet å forene (1) psykologiens autonomi som en vitenskap som er vesensforskjellig fra fysikken, med (2) våre ontologiske forpliktelser som medfører at det mentale er sterkt supervenient på det fysiske. Vår antagelse om realisme (3) genererer imidlertid en spenning mellom (1) og (2): Denne antagelsen krever nemlig at vi må finne en konkret plass for det mentale i en verden som er grunnleggende fysisk. Vi kan ikke lenger bare snakke om mentale beskrivelser, vi må også snakke om mentale *egenskaper* i den *fysiske* verden. Dette knytter vår problemstilling til den post-kartesianske varianten av sinn-legeme problemet, som Kim diagnostiserer på en treffende måte:

Through the 1970s and 1980s and down to this day, the mind-body problem—*our* mind body-problem—has been that of finding a place for the mind in a world that is fundamentally physical. The shared project of the majority of those who have worked on the mind-body problem over the past few decades has been to find a way of accommodating the mental within a principled physicalist scheme, while at the same time preserving it as something distinctive—that is, without losing what we value, or find special, in our nature as creatures with minds. (MPW 2.)

Mentale egenskapers metafysiske status i forhold til fysiske egenskaper utgjør selvsagt et kontroversielt og omfattende tema. Jeg tror likevel at kjernebekymringen til mange av de som ønsker å stå fast ved en autonom psykologi kan oppsummeres i et slags ”dilemma:” Vi må finne en plass for en ikke-fysisk vitenskap i en fysisk verden, ved å føre det mentale tilbake til denne fysiske verden. Men hva slags metafysisk relasjon gjelder mellom mentale og fysiske egenskaper? Hvis denne relasjonen blir for svak er faren at mentale egenskaper ikke får en reell plass i verden, og at psykologi dermed ikke er en vitenskap *om* verden. Vi står da ovenfor alternativene instrumentalisme eller eliminativisme. Dvs. enten er mentale begreper bare nyttige praktiske forklaringsverktøy, som det er usikkert om refererer til noe i verden, eller så finnes det ikke mentale egenskaper, og intensjonal psykologi er en *usann* teori, som bør erstattes med f.eks. moden nevrovitenskap.

På den andre siden vet vi at *hvis* det finnes mentale egenskaper i verden, så må relasjonen mellom mentale og fysiske egenskaper være minst like sterk som sterk superveniens. Grunnen til at dette er problematisk er at superveniens tilsier at disse mentale egenskapene er bestemt eller *fiksert* av fysiske egenskaper. De fysiske egenskapene er nødvendigvis tilstrekkelige for de mentale egenskapene: Gitt naturlovene og et objekts fysiske egenskaper har ikke objektets mentale egenskaper noe annet valg enn å instansieres. Fiksering innebærer at det fysiske “holder det mentale fast.” Hvis vi vil være litt mer konkrete, kan vi si

at denne fikseringen skaper et kompleks av interrelaterte problemer for psykologiens status som autonom: Gitt at fysiske egenskaper er tilstrekkelige for mentale egenskaper, hvordan kan de mentale egenskapene fungere som årsaker i tillegg til de fysiske egenskapene? Dvs. er det ikke egentlig en kompleks fysisk-biologisk egenskap som forårsaker handlingen, og ikke mitt ønske om å utføre den? Hva mer er; innebærer ikke fikseringen at vi kan forutsi de mentale egenskapenes instansieringer ut fra en rent fysisk eller biologisk teori? Kan vi derfor i prinsippet erstatte den psykologiske teorien med en fysisk eller biologisk teori? Er mentale egenskaper redusible til biologiske egenskaper?

Slike spørsmål er ofte blitt stilt både av filosofer og vitenskapsmenn. De synes å bunne i følgende bekymring: Fikseringen (sterk superveniens) gir en så tett metafysisk relasjon mellom det mentale og det fysiske at den truer med å bringe psykologien innenfor fysikkens forklaringsmessige rekkevidde. Det ser ut til at man kunne utlede psykologiske beskrivelser fra fysiske, eller erstatte psykologisk teori med fysisk teori. På denne måten truer fiksering psykologiens autonomi. Det er dette hornet av ”dilemmaet” som behandles i denne oppgaven.

Det vi trenger for å løse problemet er altså en metafysisk relasjon mellom mentale og fysiske egenskaper, som er sterk nok til å sikre at mentale egenskaper har en konkret plass i verden, og gjør en reell kausal forskjell. Dermed unngås instrumentalisme og eliminativisme. På den andre siden må den ikke være så sterk at egenskapene kommer innenfor fysikkens forklaringsmessige rekkevidde, dvs. den må unngå reduksjonisme. Emergens ser ved første øyekast ut til å være en slik relasjon: Kort sagt er emergente egenskaper makroegenskaper som bare bestemte typer objekter har. Egenskapene er nye i den forstand at de representerer kausale krefter som objektets deler ikke har, og de kan ikke forklares ut fra viten om objektets delnivå. *Det at* de emergente egenskapene instansieres når de gjør er et såkalt “brute fact,” dvs. et fundamentalt trekk ved verden som ikke tillater ytterligere forklaring. Emergentisme kan altså tilsynelatende kombinere autonomi med en tett metafysisk avhengighetsrelasjon mellom mentale og fysiske egenskaper. Dette skyldes nettopp at fysikk i følge emergentisten ikke kan gi et fullstendig svar på spørsmål som ”*hvorfor* har menneskehjernen denne bemerkelsesverdige tenkeevnen?”

Jeg vil i denne oppgaven *undersøke* om emergentisme, *hvis* det er en empirisk sann doktrine, virkelig kan sies å gi en koherent løsning av problemet. Dette må derfor ikke forveksles med et *forsvar* av emergentisme. Det skal nemlig vise seg at emergentismens sannhetsgehalt avhenger bl.a. av kausalrelasjonens natur og av hvilke metafysiske forutsetninger som vitenskapelig praksis faktisk gjør. M.a.o. er det et delvis empirisk

problem, og som sådan et problem jeg ikke tar endelig stilling til.

Det vil bemerkes at jeg godtar svært mange av Kims ontologiske og realistiske forutsetninger. Det er mulig å betvile flere av disse forutsetningene. Når jeg i denne oppgaven forutsetter dem, er det primært for å undersøke i hvilken grad det er mulig å møte Kims argumenter direkte, eller så å si på Kims hjemmebane. Men jeg håper det vil fremgå at både forutsetningene og Kims argumenter har betydelig intuitiv appell. De fortjener å taes på alvor.

Den mest sentrale utfordringen for emergentisten skal vise seg å være argumenter, fra Kim m.fl., som tilsier at emergente egenskaper ikke kan ha kausale krefter. Forpliktelsen til at emergente egenskaper kan være *årsaker* er nemlig helt sentral for emergentisten. (5.4) Jeg vil imidlertid forsøke å argumentere for at denne kritikken – gitt en rekke nye metafysiske forutsetninger – ikke er holdbar.

Til slutt i denne innledningen vil jeg komme med noen bemerkninger om problemstillingens sentrale plass i moderne filosofi: For det kan selvsagt synes som om emergentisme kun utgjør et perifert og snevert område i bevissthetsfilosofi og metafysikk. Da må det huskes at emergentisme og spørsmålet om autonomi har klar relevans for den generelle konflikten mellom reduktiv og antireduktiv fysikalisme. En viktig tese i mange antireduksjonistiske posisjoner er nemlig nettopp at fysikere ikke har *monopol* på å beskrive verden. Selv om verden er grunnleggende fysisk, så kan det finnes ikke-fysiske vitenskaper om den: Vi har allerede sett dette i Davidsons anomale monisme. Videre beskriver Robert Van Gulick antireduktiv materialisme som kombinasjonen av ontologisk materialisme (dvs. monisme) med teoretisk dualisme eller pluralisme.<sup>13</sup> Likeledes benekter Jerry Fodor slutningen fra at fysikk er en generell vitenskap til at den er den eneste vitenskap.<sup>14</sup> (Derfor kan problemstillingen kanskje også ha implikasjoner for forholdet mellom fysikk og spesialvitenskaper generelt.) Kim hevder faktisk at moderne antireduktiv materialisme er en *form for* emergentisme, og følgelig at: "[...] we have been under the reign of emergentism since the early 1970s."<sup>15</sup> Hvis dette er riktig, bør en diskusjon av emergentisme kunne lære oss noe mer generelt om antireduksjonisme.

Problemstillingen oppstår i skjæringsfeltet mellom metafysikk, bevissthetsfilosofi og vitenskapsfilosofi, og kan berøre alle disse disiplinene. Men det som gjør den mest aktuell, er kanskje at den ikke bare handler om fysikk og psykologi som akademiske disipliner, men

---

<sup>13</sup> Van Gulick (1992) s.158.

<sup>14</sup> Fodor (1974) s.127-128.

<sup>15</sup> Kim (1999) s.5.

også om fysikk og psykologi som to radikalt forskjellige perspektiver vi inntar på oss selv i hverdagen.

## 0.2 Analytisk innholdsfortegnelse

Jeg gir her en kortfattet oversikt over argumentasjonsstrukturen i oppgaven. Etter beskrivelsen av det som skjer, følger henvisning til hvilken seksjon det er snakk om i parentes. Ikke alle seksjoner er nevnt her.

**Del I:** Problemet for psykologiens status formuleres ut fra tre forutsetninger. **Kap. 1:** Psykologi og fysikk karakteriseres ut fra deres karakteristiske begreper. Psykologiske beskrivelser benytter seg i motsetning til fysiske beskrivelser av mentale verb, som er knyttet til intensjonalitet. (1.2) Psykologi og fysikk er vesensforskjellige, fordi psykologiske beskrivelser alltid forutsetter en bakgrunn av intensjonalitet, mens fysikk forutsetter lovmessige og strenge sammenhenger. De to vitenskapene er uttrykk for forskjellige erkjennelsesinteresser, og fungerer konseptuelt autonomt i forhold til hverandre. (1.3) Til tross for forskjellen er begge i en forstand uunværlige. (1.4) **Kap. 2:** Minimal fysikalisme krever (i) at alle ting har fysiske egenskaper, (ii) ikke alle ting har mentale egenskaper og (iii) mentale egenskaper avhenger av fysiske egenskaper. Svak og global superveniens oppfyller ikke disse kravene. (2.2) Sterk superveniens samt en avhengighetsrelasjon tilfredsstiller kravene. En konsekvens av minimal fysikalisme er at fysiske egenskaper nødvendigvis er tilstrekkelige for mentale egenskaper, dvs. fiksering. (2.3) Innenfor en ontologi der verden er ordnet i nivåer etter del-hele relasjonen kan minimal fysikalisme defineres som mereologisk superveniens. (2.4) **Kap. 3:** Realisme: Dersom psykologien skal ha realitet må dens begreper føres tilbake til trekk ved verden. Dette kan skje hvis den refererer til mentale egenskaper som er årsaker til handlinger. Davidson's alternative syn på dette kritiseres. (3.2) **Kap. 4:** Det er en konflikt mellom realismen, minimal fysikalisme og vesensforskjelle mellom psykologi og fysikk: Superveniens truer med å bringe det mentale innenfor fysikkens forklaringsrekkevidde, og i så fall er ikke psykologien autonom. (4.2) Emergentisme – at mentale egenskaper er emergente egenskaper – ser ut til å kunne løse konflikten. (4.3)

**Del II:** Emergentisme. **Kap.5:** Epistemologisk sett må emergente egenskaper være ikke-deduserbare fra fysiske egenskaper. (5.2) Ontologisk sett må de ha nye kausale krefter. (5.3) Emergentister er forpliktet til at de superveniente emergente egenskapene kan forårsake

egenskaper på det subveniente nivået. ("Downward causation.") (5.4) Emergensbegrepet formuleres. (5.5)

**Del III:** Kims emergentisme-kritikk diskuteres. **Kap 6:** Emergentister trues epistemologisk av Kims lokale reduksjonisme, og ontologisk av hans argumenter mot "downward causation." (6.3) Ved å gå inn for en sterk form for egenskapsdualisme kan de unngå den epistemologiske kritikken. Den ontologiske kritikken blir viktigst: Hvis psykologien skal være autonom, må mentale egenskaper kunne utøve "downward causation." (6.4) **Kap 7:** Kims superveniensargument: Hvis mentale egenskaper er irreducibile til fysiske egenskaper, så kan de ikke utøve "downward causation." (7.2) Argumentet bygger på bl.a. at det fysiske er kausalt lukket. (7.3) Prinsippet om det fysiskes kausale lukkethet kritiseres. (7.4-7.5) Det asymmetribaserte argumentet: Synkron refleksiv "downward causation" er umulig fordi det er umulig å forårsake sin egen betingelse. (7.7) **Kap 8:** For å unngå Kims argumenter kan man forkaste det fysiskes kausale lukkethet og bygge en arbeidsdelingsmodell for "downward causation." (8.1) Ifølge Roger Sperry interfererer ikke "downward causation" med mikrostrukturen. (8.2) Tvert imot er det en slags arbeidsdeling mellom mentale og fysiske egenskaper. (8.3) Problemer med modellen. Van Gulicks selektiv aktiveringsmodell for "downward causation". (8.4) Innebærer modellen en platonisk reifisering av mønstrene delene er organiserte i? Svaret synes å avhenge av bl.a. intuisjoners holdbarhet og vitenskapelig praksis. (8.5) **Kap 9:** Avsluttende refleksjoner om oppgavens problem og antireduksjonisme.

# **Del I: Problemet for psykologiens status**

## **1 Psykologi og fysikk er vesensforskjellige disipliner**

### **1.1 Innledning**

Jeg vil i dette kapittelet argumentere for at psykologi og fysikk er vesensforskjellige hva gjelder de begreper de benytter seg av, og anvendelsesbetingelsene for disse. Denne forskjellen kommer godt frem i Davidsons klassiske artikkel "Mental Events" (1970), som min diskusjon tar utgangspunkt i. Avslutningsvis argumenterer jeg kort for at til tross for denne forskjellen mellom fysikk og psykologi, så synes begge i en viss forstand å være uunværlige.

### **1.2 Karakterisering av psykologi og fysikk**

Jeg forutsetter her at vitenskaper kjennetegnes ved deres karakteristiske begreper og betingelsene for anvendelsen av disse. Eksempler på vitenskapers karakteristiske begreper er at fysikk benytter seg av begreper som "ladning," "masse," "kraft," kjemi bruker bl.a. "løselig," "sur," mens psykologi typisk refererer til f.eks. "ønsker" og "tanker." Bruken av disse begrepene er knyttet til de erkjennelsesinteresser denne bruken er uttrykk for: Når vi eksempelvis bruker fysiske begreper, er vi typisk sett ute etter maksimalt generelle og eksakte sammenhenger mellom begrepene. Newtons bevegelseslover er således forutsigelsesverktøy som postulerer strenge lovmessige og generelle sammenhenger mellom begreper som "masse" og "akselerasjon." En anvendelsesbetingelse for fysiske begreper er derfor at man antar at sammenhenger mellom begrepene prinsipielt kan skjerpes til generelle lover.

Bruken av psykologiske begreper er derimot knyttet til ønsket om å forstå oss selv og andre som handlende og tenkende vesener, og om å samhandle ut fra dette. Anvendelsesbetingelsen for disse begreper skal derfor vise seg å være at man antar at personer stort sett har sanne oppfatninger og stort sett er rasjonelle.

Hovedpoenget med denne karakteriseringen av vitenskaper er altså at forskjellige erkjennelsesinteresser gir forskjellige anvendelsesbetingelser for begrepene, og at man på denne måten kan ende opp med vitenskaper som er vesensforskjellige begrepslige systemer.

Denne språklige innfallsporten til vitenskaper er ikke uforenelig med å se på vitenskaper som praksiser, det vil grovt sett si som delte oppfatninger, interesser og handlinger (f.eks. eksperimenter) hos en sosial gruppering. Tvert imot er vitenskaper å

betrakte som språklige praksiser. At det er den språklige dimensjonen ved praksisen som det fokuseres på her, utelukker heller ikke andre dimensjoner ved praksisen.

Davidsons diskusjon i "Mental Events" tar utgangspunkt i en metafysikk av hendelser ("events"), som han beskriver og eksemplifiserer som "unrepeatable, dated *individuals* such as the particular eruption of a volcano, the (first) birth or death of a person, [...]"<sup>16</sup> Jeg kommer imidlertid hovedsakelig til å skrive innenfor en metafysikk av objekter og egenskapsinstansieringer, der hendelser *utgjøres* av et objekts instansiering av en egenskap ved et bestemt tidspunkt. Litt mer formelt kan vi for et objekt *x*, en egenskap *P* og et tidspunkt *t* snakke om en hendelse *H* som *er* instansieringen *P(x)* ved *t*.<sup>17</sup>

Davidsons forsøk på å karakterisere det mentale versus det fysiske starter fra et forslag om at hendelser er mentale hvis de kan gies en mental *beskrivelse*: "[...] an event is physical if it is describable in a purely physical vocabulary, mental if describable in mental terms."<sup>18</sup> Dette harmonerer med den språklige vinklingen på vitenskaper. Davidson fortsetter fra dette utgangspunktet, og definerer en mental hendelsesbeskrivelse som en hendelsesbeskrivelse som inneholder et mentalt verb essensielt:

Let us call a description of the form 'the event that is *M*' or an open sentence of the form 'event *x* is *M*' a *mental description* or a *mental open sentence* if and only if the expression that replaces '*M*' contains at least one mental verb essentially.<sup>19</sup>

(Essensielt betyr her at det mentale verbet ikke kan elimineres ved substitusjon med en ekvivalent setning *uten* et slikt verb.)<sup>20</sup> Mentale verb er verb som kan inngå i proposisjonale attityder, eller påstandsholdninger: "We may call those verbs mental that express propositional attitudes like believing, intending, desiring, [...]"<sup>21</sup> Dermed får Davidson en definisjon eller karakterisering av mentale og fysiske hendelser:

Now we may say that an event is mental if and only if it has a mental description, or (the description operator not being primitive) if there is a mental open sentence true of that event alone. Physical events are those picked out by descriptions or open sentences that contain only the physical vocabulary essentially.<sup>22</sup>

To vesentlige konsekvenser følger av den karakteriseringen Davidson har valgt å gi det mentale: For det første, som Davidson selv bemerker: "[...] the distinguishing feature of

---

<sup>16</sup> Davidson (1970) s.209, min kursivering. (Eventuelle tilføyelser i klammeparanteser ([...]) i sitater er mine.)

<sup>17</sup> Mer om dette følger i (3.2).

<sup>18</sup> Davidson (1970) s.210

<sup>19</sup> Ibid. s.211. I det som følger kommer jeg til å fokusere på bestemte hendelsesbeskrivelser, og ikke åpne setninger.

<sup>20</sup> Davidson (1970) s.211.

<sup>21</sup> Ibid. s.210

<sup>22</sup> Ibid. s.211. Davidson vedkjenner seg (ibid.) at han her ikke har gitt noen omfattende beskrivelse av det fysiske vokabularet.



the mental is not that it is private, subjective, or immaterial [these are classical distinguishing features of the mental], but that it exhibits what Brentano called intentionality.”<sup>23</sup> Hvorfor følger dette? Grunnen er at en fruktbar og vanlig måte å forstå intensjonalitet på, er som det å være *rettet mot* noe eller å *forholde seg til* noe. Mentale verb uttrykker nettopp måter personer kan forholde seg til en påstand eller proposisjon p, dvs. de kan uttrykke påstandsholdninger: Hvis p f.eks. er ”Ole består eksamen,” så kan Ole *frykte* at  $\neg p$ , og Oles mor kan *tro* at p.<sup>24</sup>

For det andre omfatter den begrepslige praksis vi har skissert psykologi i nokså vid forstand: For både den psykologi vi alle bruker for å forstå oss selv og andre i hverdagen (”folkepsykologi”), og deler av akademisk psykologi – i det minste store deler av psykoanalytiske og praktisk-psykiatriske tradisjoner – benytter seg i høy grad av påstandsholdninger. Dette er tydelig hvis man f.eks. tenker på den viktige rolle forklaringer av handlinger med oppfatninger og ønsker spiller. Hvorfor leser Ole så flittig? Jo, han *ønsker* å bestå eksamen, han *tror* at dette er mulig for ham og han *tror* at det å lese flittig kan føre til at han består eksamen. De nevnte tradisjonene i akademisk psykologi har formodentlig en mer subtil historie å fortelle (f.eks. om hvilke faktorer som formet Oles oppfatninger, eller om eventuelle underbevisste ønsker han har etc.), men også denne psykologien anvender påstandsholdninger på en ikke-eliminerbar måte.<sup>25</sup> Det er også verdt å merke seg at sosialvitenskaper som økonomi eller statsvitenskap i høy grad betrakter personer som rasjonelle aktører, og følgelig forutsetter intensjonalitet.

Vi har altså karakterisert psykologi og fysikk som *begrepslige* praksiser, på en foreløbig metafysisk relativt nøytral måte. ”Nøytral” fordi vi ikke har forutsatt noe om egenskapers metafysiske status, eller om det det ontologiske forholdet mellom fysiske og mentale egenskaper. Videre har vi ved å akseptere den davidsonianske intensjonale karakteriseringen av det mentale fanget et sentralt, om ikke også *essensielt* aspekt ved flere former for psykologi.

---

<sup>23</sup> Davidson (1970) s.211.

<sup>24</sup> Det er derfor ikke tilfeldig at mentale verb ofte, som Davidson bemerker, brukes til å referere til personer. (Ibid. s.211) Tvert imot kan det mentales intensjonalitet forklare en viss asymmetri mellom psykologiske og fysiske begrepslige praksiser: Fysiske termer brukes om alle fysiske objekter, mens mentale termer (oftest) brukes om en begrenset klasse fysiske objekter, nemlig personer. Svarende til denne begrepslige asymmetrien er som vi skal se (2.2) en metafysisk asymmetri mellom mentale og fysiske egenskaper: Alle objekter har fysiske egenskaper, mens bare noen objekter har mentale egenskaper.

<sup>25</sup> Her må jeg åpne for unntak for visse former for biologisk psykologi, og annen ikke-intensjonal akademisk psykologi. F.eks. tar Paul og Patricia Churchland til orde for et alternativ til psykologi *a lá* påstandsholdninger. P.M. Churchland foreslår at folkepsykologi faktisk er en empirisk *usann teori*, som bør elimineres (Churchland (1996) s.8) til fordel for en teori om transformasjoner mellom aktiveringsvektorer, som er familier av nerveceller som aktiveres. (Ibid. s.11.) Den radikale ideen er altså at det *ikke finnes* ønsker og oppfatninger, men bare nerveceller i interaksjon, og at den eneste sanne psykologi handler om sistnevnte. Men slike former for psykologi vil ikke pretendere å være autonom i forhold til fysikk, så de vil heller ikke være relevante for denne oppgaven.

### 1.3 Forskjellen mellom psykologi og fysikk

Davidson kommer i "Mental Events" frem til at det er en vesentlig forskjell mellom mentale og fysiske beskrivelser. I stedet for å følge Davidsons argumentasjon slavisk og kronologisk, beskriver jeg tre måter man kan innse denne forskjellen på når man reflekterer over denne tematikken og/eller leser "Mental Events." Dette svarer til tre suksessive argumentative trinn man kan gjennomgå for å oppnå en gradvis mer begrunnet oppfatning av forskjellen mellom mentale og fysiske begrepslige praksiser.

Trinn (1): Ved observasjon av hvordan mentale begreper fungerer kan man finne ting som tyder på at det er en forskjell mellom psykologiske og fysiske beskrivelser. Man kan nemlig se at mentale begrepslige praksiser er preget av en *holisme*, som avgrenser dem fra praksiser som utelukkende benytter seg av behavioristiske og/eller fysiske termer. Davidson ber oss tenke på at definitorisk behaviorisme slår feil, og at mentale termer derfor ikke lar seg definere utelukkende ved termer fra et vokabular om adferd.<sup>26</sup> Hans eget eksempel på en mislykket reduksjon illustrerer dette:

Suppose we try to say, not using any mental concepts, what it is for a man to believe there is life on Mars. One line we could take is this: when a certain sound is produced in the man's presence ('Is there life on Mars?') he produces another ('Yes'). But of course this shows he believes there is life on Mars only if he understands English, his production of the sound was intentional, and was a response to the sounds as meaning something in English; and so on. For each discovered deficiency, we add a new proviso. Yet no matter how we patch and fit the non-mental conditions, we always find the need for an additional condition (provided he *notices*, *understands*, etc.) that is mental in character.<sup>27</sup>

Den underliggende strukturen i slike velkjente antireduktive argumenter er ikke å komme opp med sære scenarier der den fysisk-behavioristiske definisjonen ikke holder. Tvert imot består argumentene i å påpeke at det mentale er preget av en holisme: Oppfatninger forutsetter f.eks. alltid andre oppfatninger, slik sitatet viser. Davidson sier at det er et mønster i de mislykkete reduksjonene ("pattern of failure"), og derfor:

Beliefs and desires issue in behaviour only as modified and mediated by further beliefs and desires, attitudes and attendings, without limit. Clearly this holism of the mental realm is a clue both to the autonomy and to the anomalous character of the mental.<sup>28</sup>

Slik kan man nærmest ved observasjon innse at det mentale utgjør et hele, og at psykologiske begrepslige praksiser ikke kan reduseres *definitorisk* til fysiske/behavioristiske begrepslige praksiser. Men dette kunne tenkes å ganske enkelt skyldes et trekk ved mentale termers semantikk: Nemlig det at de henger sammen og forutsetter hverandre, uten at det

---

<sup>26</sup> Davidson (1970) s.216

<sup>27</sup> Ibid. s.217

<sup>28</sup> Ibid.

nødvendigvis av den grunn er noen fundamental begrepslig forskjell mellom dem og fysiske termer. Imidlertid har holismen – og dermed den konseptuelle kløften mellom psykologi og fysikk – en dypere forklaring. Det er kanskje denne dypere forklaringen Davidson mener at holismen er et ”clue” til.

Trinn (2): Hva skyldes så holismen; hvorfor inngår mentale beskrivelser i et hele som er avgrenset fra fysiske beskrivelser? Holismen kan forklares ut fra at praksisene har forskjellige og motstridende anvendelsesbetingelser. Det å gå fra den psykologiske praksisen til den fysiske praksisen krever derfor at man skifter regelsett.

Mulighetsbetingelsen for anvendelse av påstandsholdninger på en person er at denne anvendelsen må skje innenfor en helhetlig teori som er strukturert rundt rasjonalitet:

*Just as we cannot intelligibly assign a length to any object unless a comprehensive theory holds of objects of that sort,<sup>29</sup> we cannot intelligibly attribute any propositional attitude to an agent except within the framework of a viable theory of his beliefs, desires, intentions, and decisions.<sup>30</sup>*

I det øyeblikket vi tilskriver noen en påstandsholdning, la oss si et ønske, må vi forstå dette ønsket i sammenheng med andre ønsker, oppfatninger og desslike, og i.o.m. denne forpliktelsen har vi generert det mentales holisme. Holismen ligger i at vi ikke kan tilskrive noen én holdning uten å implisitt tilskrive ham et udefinert antall andre holdninger. Davidson sier at holdningene får sitt innhold fra sin plass i ”mønsteret,” dvs. i helheten.<sup>31</sup> Prinsippet for forklaring når vi bruker slike påstandsholdninger er rasjonalitet i vid forstand. For når vi har tilskrevet personen et hele av holdninger, kan vi ikke annet enn å forvente en stor grad av konsistens, rasjonalitet og sanne oppfatninger. Holismen følger av dette prinsippet.

At denne rasjonalitetsbetingelsen er konstituerende for den begrepslige praksis der holdningene inngår, er kanskje ikke overraskende gitt at det mentale er kjennetegnet ved intensjonalitet.

Rasjonalitetsbetingelsen for anvendelsen av psykologiske termer er Davidsons berømte velvillighetsprinsipp, eller ”Principle of Charity:” Vi må tilskrive aktører en stor grad av konsistens og kunnskap. Bare mot en slik bakgrunn kan vi diagnostisere irrasjonalitet og feil.<sup>32</sup> Det er altså ikke nok at anvendelsen av verbene som uttrykker påstandsholdninger på en person forutsetter at vi tilskriver personen en videre psykologi av oppfatninger, ønsker osv. (Dette så vi allerede ovenfor da vi konstaterte at det mentale var preget av en holisme.) Vi må

---

<sup>29</sup> Dette viser til Davidsons instruktive eksempel om begrepet ”lengde:” Idet vi anvender lengdebegrepet forutsetter vi deler av en større bakgrunnsteori. Vi må f.eks. forutsette at lengderelasjonen er transitiv. (Ibid. s.220)

<sup>30</sup> Davidson (1970) s.221, min kursivering.

<sup>31</sup> Ibid.

<sup>32</sup> Ibid. Dette prinsippet utdypes i artikler som ”On the Very Idea of a Conceptual Scheme” (Davidson (1974)).

også i vår fortolkning av henne forstå henne som stort sett rasjonell og stort sett innehavende sanne oppfatninger:

[...] when we use the concepts of belief, desire, and the rest, we must stand prepared, as the evidence accumulates, to adjust our theory in the light of considerations of overall cogency: the constitutive ideal of rationality partly controls each phase in the evolution of what must be an evolving theory.<sup>33</sup>

Davidsons karakterisering av fysikk er mindre omfattende,<sup>34</sup> men det er tydelig at han regner muligheten for å finne strenge lover som sentral i den fysiske praksis. Med *strenge* lover vil jeg forstå *unntaksløse* lover, dvs. lover uten *ceteris paribus* klausuler. En lov er *ceteris paribus* hvis den kun holder under normale omstendigheter, dvs. kun når en rekke betingelser som ikke er spesifiserte i loven er tilfredsstilt. Dette harmonerer med Davidsons syn på strenge lover, slik han uttrykker det i "Laws and cause":

Strictly lawlike statements do not contain open-ended phrases like "other things being equal", or "under normal conditions".<sup>35</sup>

Strenge lover kan ifølge Davidson bare finnes hvis lovene henter sine begreper fra en omfattende lukket teori. Hans videre påstand er at fysiske generaliseringer er av en slik art at vi – i hvert fall prinsipielt – kan forvente å kunne presisere dem til slike lover.<sup>36</sup> Det er fristende å nevne kvantifisering som et annet sentralt aspekt ved bruken av fysiske termer. (Det er muligheten til å beregne som har gjort matematisk fysikk til et slikt enormt nyttig forutsigelsesverktøy.) Beskrivelser som inneholder kvantifiseringer synes godt egnet til å uttrykke strenge sammenhenger, nettopp fordi kvantifisering muliggjør en høy grad av presisjon.

Vi kan nå slå fast at det er en prinsipiell forskjell mellom mellom fysikk (som fordrer streng lovmessighet og dermed generelle sammenhenger) og psykologi (som uungåelig forutsetter intensjonalitet, rasjonalitet og sanne oppfatninger):

There are no strict psychophysical laws because of the disparate commitments of the mental and physical schemes. It is a feature of physical reality that physical change can be explained by laws [i.e. *strict* laws] that connect it with other changes and conditions physically described. It is a feature of the mental that the attribution of mental phenomena must be responsible to the background of reasons, beliefs and intentions of the individual. There cannot be tight connections between the realms if each is to retain allegiance to its proper source of evidence.<sup>37</sup>

---

<sup>33</sup> Davidson (1970) s.223

<sup>34</sup> Han sier selv at han i den herværende artikkelen kommenterer det fysiske vokabularets natur, men at dette ikke gir et kriterium for vokabularet. (Ibid. s.211)

<sup>35</sup> Davidson (1995) s.266, se også (1970) s.208.

<sup>36</sup> Davidson (1970) s.219.

<sup>37</sup> Ibid. s.222. Det vil bemerkes at Davidson mener dette også er et generelt antireduktivt argument, ikke bare et argument mot definitorisk reduksjonisme. Hvorvidt denne *generelle* konklusjonen følger er kontroversielt, og jeg forutsetter ikke dette. Som vi skal se kan emergentisten faktisk godta bikondisjonaler som forbinder mentale og fysiske egenskaper, så lenge det at bikondisjonalene holder er et såkalt "brute fact."

Det er altså en konseptuell kløft mellom fysikk og psykologi. Deres anvendelsesbetingelser er ikke bare forskjellige, men også motstridende. (Men de står vel å merke ikke i noen *logisk* motsetning.) Å forsøke å definere begreper i den ene praksisen ut fra begreper i den andre vil være et regelbrudd, som vi har sett umiddelbart straffes med at man må gjøre tilleggsforutsetninger i den første praksisen.

At det er en slik konseptuell kløft mellom det mentale og det fysiske, er noe man i en viss forstand kan innse *a priori*: For Davidson hevder at det å akseptere muligheten av strenge psykofysiske lover ville innebære å forandre samtaleemnet ("changing the subject"), ved ikke å akseptere at det mentale vokabularet kjennetegnes av påstandsholdningene.<sup>38</sup> Kløften er altså *a priori*, i den forstand at den er gitt så snart mulighetsbetingelsene for praksisene er på plass. Men man kan selvsagt velge å snakke på alternative måter, med andre mulighetsbetingelser. En ikke-intensjonal biopsykologi er en slik mulighet.

Det er viktig å innse at dette ikke innebærer at psykologi ikke kan benytte seg av fysiske termer, dette skjer ofte. Poenget er at psykologi uungåelig vil måtte ha et betydelig intensjonalt aspekt. (Davidson er forøvrig også inneforstått med at det må finnes sanne ikke-strenge sammenhenger mellom fysikk og psykologi.)<sup>39</sup>

Alle vil selvsagt ikke enes om detaljene i Davidsons argument omkring forholdet mellom fysikk og psykologi. Eksempelvis kan uenighet forventes om hans karakteristikkk av kausalitet ut fra generalitet og lovmessighet. Dette er mindre viktig. Hans generelle karakteristikkk av psykologi som en begrepslig praksis kjennetegnet ved intensjonalitet, får på en god måte frem at psykologi er vesensforskjellig fra fysikk, under enhver rimelig karakterisering av den sistnevnte vitenskapen. Det må i det minste vektige motargumenter til for at denne hovedkonklusjonen skal forkastes, og de færreste vil vel sette mye inn på å konstruere slike argumenter. Davidson sier faktisk selv at: "That there is a categorical difference between the mental and the physical is a *commonplace*."<sup>40</sup>

Trinn (3): En pragmatisk tilnærming kan la oss se hvorfor psykologi og fysikk har forskjellige anvendelsesbetingelser, og følgelig sette den konseptuelle forskjell mellom dem i en videre sammenheng. For det synes svært rimelig at psykologi og fysikk er oppstått som følge av forskjellige praktiske hensyn eller erkjennelsesinteresser: Vi bruker psykologi til (bl.a.) å forstå og forutsi oss selv og andre, særlig med henblikk på adferd og samhandling.

---

<sup>38</sup> Davidson (1970) s.216. (Om dette antireduktive resultatet følger for psykofysiske lover generelt er usikkert, men for definitorisk reduksjon synes saken klar.)

<sup>39</sup> Davidson (1970) s.216. Han sier faktisk også at han selv *forutsetter* slike sammenhenger i (1993) s.14.

<sup>40</sup> Davidson (1970) s.223, min kursivering.

Psykologiske beskrivelser fremstiller mennesket på en slik måte at det kan evalueres ut fra normer og formål. Fysikk bruker vi derimot til f.eks. å måle temperaturen når vi har feber, bygge atombomber og finne ut når det er gunstigst å sende romskip til Mars. Fysikk er uttrykk for et ønske om å finne generelle, lovmessige og strenge sammenhenger, ikke for normativitet. Når vi setter alle andre interesser enn interessen i maksimal generalitet tilside bedriver vi *teoretisk* fysikk, ikke fysiske spesialvitenskaper som geologi etc.<sup>41</sup> Det er intuitivt plausibelt at disse praksisene skal ha forskjellige prinsipper og begreper siden de svarer til forskjellige hensyn.

Denne intuisjonen kan gies svært konkret mening, og bygges ut til et pragmatisk argument for at mentale og fysiske begreper ikke kan settes i pene og entydige reduktive sammenhenger. Et slikt argument finner vi hos Van Gulick. Han peker bl.a. på at forskjellige formål gir forskjellig konseptuell oppdeling av verden med hensyn både til hvor ”vide” eller ”trange” vitenskapens slag er, og med tanke på hvilke faktorer som er relevante for individuering.<sup>42</sup> Han oppsummerer argumentet slik:

[...] we should not expect to be able to establish exact correspondences between the concepts of our first-person self-conscious representational framework and the concepts of the physical sciences. They are too different with respect to how the representation-user causally interacts with the representations, too different with respect to the goals which the framework can be used to achieve, and too different with respect to the role played by indexical and self-referential factors in determining content. Some measure of incommensurability is what we should expect.<sup>43</sup>

En interessant metafor understreker den praktiske dimensjonen ved teoriene: Van Gulick sier at mentale og fysiske rammeverk representerer de samme spatio-temporale regionene, men setter forskjellige håndtak eller ”handles” på dem.<sup>44</sup> Hva denne metaforen betyr kommer kanskje klarest frem i en annen artikkel, der han vektlegger sin forståelse av teorier som *organiserte representasjonssystemer vi bruker til å lede våre handlinger*. Kausale lover i teorier oppretter forbindelser mellom våre representasjoner ”that mirror stable recurrent patterns in the represented world.” Dermed kan de rettlede oss i våre handlingspraksiser (og formodentlig også forutsigelsespraksiser): ”As causal laws they specifically single out the independent variables in such patterns, the levers that can and must be pushed to produce desired changes.”<sup>45</sup>

Ved å se på teorier ut fra et pragmatisk perspektiv har Van Gulick altså gitt en plausibel forklaring på hvorfor psykologi og fysikk – til tross for at de er *om* den samme

---

<sup>41</sup> Jf. Bjørn Ramberg (1999) s.612.

<sup>42</sup> Van Gulick (1992) s.168.

<sup>43</sup> Ibid. s.170-171.

<sup>44</sup> Ibid. s.168.

<sup>45</sup> Van Gulick (1993) s.246.

verden – er forskjellige i sitt vesen. Det han så langt ikke har gitt er en *metafysisk* redegjørelse for hvilken relasjon til verden som gjør at begge disse representasjonssystemene *virker så godt*. Jeg vil våge påstanden om at vi nå er ved grensen for det ren språkfilosofi og pragmatisme kan (og kanskje vil) forklare.<sup>46</sup> Dersom man beveger seg inn i det filosofiske landskapet som venter, må man være forberedt på å møte rent *ontologiske* problemer.

#### 1.4 Avsluttende kommentarer

Det er etablert at vi har to vesensforskjellige begrepslige praksiser (psykologi og fysikk) og at det finnes gode pragmatiske forklaringer av denne forskjellen. Det er videre klart at begge i en forstand er uunnværlige: Det skal svært vektige grunner til for å forkaste dem, hvis det da i det hele tatt er psykologisk mulig for oss å tre ut av praksisene ved å slutte å bruke deres karakteristiske begreper:

Fysisk teori er i det store og det hele så godt empirisk bekreftet, og har gitt opphav til så mange vellykkede forutsigelser og teknologiske nyvinninger, at den er blitt integrert som en uunnværlig del av manges verdensbilde. Dessuten kan det hevdes at en eller annen form for fysisk vokabular – forstått som uttrykk for interesse i strenge generelle sammenhenger – er noe vi er biologisk disponerte for å utvikle og bruke: At mennesket så å si er et *induktivt dyr* fra naturens side, og uungåelig må søke generelle sammenhenger, synes implisitt allerede i Humes kausalitetsanalyse.<sup>47</sup> Hos Davidson er denne innsikten blitt mer eksplisitt: Vi er ”inducers from birth”, og *må* finne generelle sammenhenger for å overleve. Vår idé om kausale sammenhenger bygger på dette.<sup>48</sup> Fysikk er et uttrykk for vår interesse i lovmessigheter: ”Our concept of a physical object is the concept of an object whose changes are governed by law.”<sup>49</sup> Gitt dette ser det ut til at vi ikke *kan* tre ut av den fysiske begrepslige praksisen.

På den andre siden bruker vi (folke)psykologi ustanselig. Selv om vi selvsagt av og til tar feil, så er denne bruken så inngrodd at vi vanligvis ikke reflekterer over teorien så lenge

---

<sup>46</sup> Et eksempel på et slikt syn på språkfilosofi finner vi i kap. VI av Richard Rorty’s ”Philosophy and the Mirror of Nature”: Språkfilosofi bør forbli ”pure” – dvs. kun omhandle hvordan språket virker – og ikke vikle seg inn i spørsmål omkring språkets implikasjoner for andre, spesielt epistemologiske, filosofiske problemer. (Rorty (1980) s.267.)

<sup>47</sup> ”Custom, then, is the great guide of human life. [...] Without the influence of custom, we should be entirely ignorant of every matter of fact beyond what is immediately present to the memory and senses.” (Hume (1927) s.45=*Enquiry* Sect.V, Part I.) Med ”custom” mener Hume at vi forventer at (f.eks.) biljardkuler vil sette hverandre i bevegelse også i fremtiden. Mennesket søker altså generelle kausale sammenhenger, ikke bare viten om enkelte korellasjoner av hendelser. Videre sier Hume at dette er et ”instinct” som ”nature” har implantert i oss. (Ibid. s.56=*Enquiry* Sect.V, Part II.)

<sup>48</sup> Davidson (1995) s.275.

<sup>49</sup> Ibid. s.274. Jf. Ramberg (1999) s.610-611.

den fungerer.<sup>50</sup> En av folkepsykologiens fremste proponenter, Fodor, sier det slik:

”Commonsense psychology works so well it disappears.”<sup>51</sup>

Davidsons perspektiv på psykologi og fysikk som begrepslige praksiser synes å gi oss alt vi ønsker oss: For det synes å love at vi kan kombinere det at psykologi og fysikk er vesensforskjellige, autonome og uunnværlige med våre ontologiske forpliktelser ovenfor fysikken. For i et berømt argument mener Davidson å utlede at alle mentale hendelser er token-identiske med fysiske hendelser: De må nemlig ha en fysisk beskrivelse, i tillegg til den mentale beskrivelsen. (Det overraskende med denne utledningen av monisme, er at et av premissene er at psykologien er *autonom*, i den forstand at det ikke finnes noen strenge psykofysiske lover.)<sup>52</sup> Alt ser ut til å være vel. Imidlertid skal vi se at det ikke er så enkelt: Våre ontologiske forpliktelser kommer i neste kapittel til å generere problemer for psykologiens autonomi.

## 2 Minimal fysikalisme

### 2.1 Innledning

Målet i dette kapittelet er å definere “minimal fysikalisme,” et begrep som ble innført i innledningen.<sup>53</sup> Vi skal altså finne en metafysiske relasjon mellom det mentale og det fysiske som er svak nok til at den kan aksepteres av både reduksjonister og antireduksjonister, men som samtidig tar hensyn til det naturvitenskapelige verdensbildet. Hvis ikke annet er indikert, vil jeg følge Kim i å ha en vid forståelse av hva som er fysisk. (MPW 113.) Når jeg snakker om fysiske egenskaper i forbindelse med mentale egenskaper, vil det følgelig ofte dreie seg om biologiske (nevrofysiologiske) egenskaper.

Etter at Davidson innførte superveniensbegrepet i moderne bevissthetsfilosofisk debatt, har mye av diskusjonen omkring forholdet mellom det mentale og det fysiske foregått ut fra superveniens.<sup>54</sup> I tråd med dette vil jeg her lete etter en superveniensrelasjon som kan definere minimal fysikalisme.

---

<sup>50</sup> Hvorvidt folkepsykologi er en *teori* – og dermed hvorvidt man kan stille spørsmål ved dens legitimitet, slik Churchland forutsetter – er selvsagt kontroversielt. (Se f.eks. Churchland (1988) s.58-59, (1996) s.10-12.) Noen vil hevde at folkepsykologi snarere er en naturlig praksis vi – *qua* de biologiske vesener vi er – tar del i. Min diskusjon omhandler imidlertid *ekspliserte* folkepsykologiske prinsipper, forklaringer og predikasjoner, og disse *kan* betraktes som en teori (om den nevnte naturlige praksisen). Som sådan kan den beskrives som en metateori, som legger føringer på intensjonal-psykologisk forskning.

<sup>51</sup> Fodor (1987) s.3, min kursivering.

<sup>52</sup> Davidson (1970) s.224.

<sup>53</sup> Begrepet finnes hos Kim: MPW 15 og Kim (1998b) s.12. Selv om min motivasjon og begrunnelse av minimal fysikalisme i noen grad avviker fra Kims, er mitt begrep “minimal fysikalisme” forenelig med hans.

<sup>54</sup> Davidson (1970) s.214.



Jeg vil først sette frem tre suksessivt sterkere ontologiske krav til minimal fysikalisme. Derneft bygger jeg opp superveniensbegrepet fra de svakeste mulige metafysiske forutsetningene, til jeg har funnet et begrep som tilfredsstillar kravene til minimal fysikalisme. Ved å gå frem på denne måten unngår jeg å inkorporere for sterke forutsetninger i minimal fysikalisme. Jeg begynner derfor med å undersøke om såkalt svak eller global superveniens tilfredsstillar disse kravene. Det skal vise seg at disse superveniensrelasjonene er for svake, og at vi trengjer såkalt *sterk* superveniens samt en avhengighetsrelasjon i definisjonen vår. Jeg beskriver så ontologien av en verden ordnet i nivåer etter del-hele relasjonen, der minimal fysikalisme kan defineres som såkalt mereologisk superveniens. Underveis vil det fremgå at minimal fysikalisme innebærer at det mentale bestemmes eller fikseres av det fysiske. Dette kapittelet om minimal fysikalisme er omfattende, fordi det setter opp de ontologiske forutsetningene for resten av oppgaven.

## 2.2 Kravene til minimal fysikalisme. Svak og global superveniens

Hvilke krav må vi stille til en ontologi som ikke skal være substansdualistisk (eller substans-pluralistisk), og som skal ta vare på det naturvitenskapelige verdensbildet? Substansdualisme forstås her rent negativt: Vi vil ikke ha en ontologi med to vesensforskjellige typer av ting. Dette kan utelukkes *a priori*, men vi har også empiriske grunner til å tro at den ene substansen som eksisterer er den *fysiske* substansen. Dermed er vi forpliktet til (1) *Monisme*: Det finnes bare en substans, og denne substansen er den fysiske materie. Innenfor en ontologi av objekter og egenskaper betyr dette at alle objekter har fysiske egenskaper.<sup>55</sup> (2) *Asymmetri*: Imidlertid forteller erfaringen oss at alle objekter ikke omvendt har mentale egenskaper.<sup>56</sup> (3) *Avhengighet*: Mentale egenskapers eksistens avhenger i en forstand av eksistensen av fysiske egenskaper, men det omvendte ikke er tilfellet.<sup>57</sup>

---

<sup>55</sup> Hvis ikke annet er indikert forstår jeg heretter monisme og dualisme som gjeldende substanser, og ikke egenskaper eller teorier. Det vil bemerkes at jeg nettopp vekslet mellom to betydninger av substans (hhv. "materie" eller eng. "stuff" og "type ting"). For minimale fysikalister er monisme i begge betydningene en forpliktelse. Betydningene er selvsagt ikke uavhengige, men jeg kommer til å fokusere på den siste betydningen.

<sup>56</sup> Hvis man, som Davidson, opererer med et svakt begrep "mentalt" der noe er mentalt hvis det kan beskrives med psykologiske termer, blir *alle* hendelser eller egenskapsinstansieringer mentale: For hendelser kan beskrives mentalt ved å si at de er samtidige med en mental hendelse. (Davidson (1970) s.211-12.) For hendelser som er fjernt i tid fra mentale hendelser (f.eks. "Big Bang"), vil jeg foreslå at man likeledes kan spesifisere tidsintervallet mellom hendelsen og den mentale hendelsen i beskrivelsen. Men det er klart at slike hendelser bare er mentale i en aksidensiell forstand. Dette (og kravet om asymmetri) betyr imidlertid ikke at det aldri er relevant å gi metafysisk sett fysiske objekter mentale beskrivelser. Mount Everest er et fysisk objekt, men kan gies en mental beskrivelse: For vi kan snakke om det som "det fjellet Mallory ønsket å bestige."

<sup>57</sup> Kim nevner ofte "bestemmelse" eller "determination" i samme åndedrag som "dependence". (F.eks. MPW 11.) Men i denne sammenhengen fokuserer jeg på avhengighet. Bestemmelse eller fiksering følger som vi skal se som et slags korollar fra minimal fysikalisme.

Etter at Davidson i “Mental Events” (1970) – nærmest *en passant* – innførte superveniensbegrepet i moderne bevissthetsfilosofi har det, som nevnt, vært vanlig å bruke superveniens til å formulere den metafysiske avhengighetsrelasjonen mellom det mentale og det fysiske. Davidsons eget superveniensbegrep er en variant av såkalt ”svak superveniens.”<sup>58</sup> La oss derfor undersøke om svak superveniens er en sterk nok relasjon til å tilfredsstille våre tre krav:

Kim formulerer svak superveniens som en relasjon mellom to mengder av egenskaper A og B, som vi kan tenke på som inneholdende henholdsvis mentale og fysiske egenskaper. At de mentale egenskapene i A er svakt superveniente på de fysiske egenskapene i B kan da gies to ekvivalente definisjoner:<sup>59</sup>

*Weak covariance I:* No possible world contains things, x and y, such that x and y are indiscernible in respect of properties in B (“B-indiscernible”) and yet discernible in respect of properties in A (“A-discernible”).

*Weak covariance II:* Necessarily, if anything has property F in A, there exists a property G in B such that the thing has G, and everything that has G has F.<sup>60</sup>

I denne sammenhengen behøver vi ikke anta at “necessarily” innebærer noe sterkere enn nomologisk nødvendighet: ”Påstanden p er nomologisk nødvendig” betyr at p er nødvendig gitt naturlovene slik de faktisk er, hvilket igjen vil si at p er sann i alle mulige verdener som har samme lover som den faktiske verden.

Svak superveniens gir altså en systematisk kovariasjon mellom et objekts mentale og fysiske egenskapsinstansieringer. Vi må imidlertid merke oss i hvilken grad denne kovariasjonen er nødvendig: Definisjon II forteller oss (i) at det er *nødvendig* at hvis et objekt x har en mental egenskap F, så har det også en fysisk egenskap G, og (ii) at G er tilstrekkelig for F. Men (ii) er *ikke* nødvendig, og definisjon I uttaler seg likeledes bare om forhold i *én* mulig verden. Det at F instansieres når G instansieres er altså *kontingent*. Konsekvensen av dette blir at svak superveniens er forenelig med at forholdet mellom den mentale egenskapen F og den fysiske egenskapen G kan være annerledes i andre mulige verdener. Som Kim sier:

*[T]he fact that mentality is distributed in a certain way in one world has absolutely no effect on how it might be distributed in another world. Intra-world consistency of the distribution of mental properties relative to the distribution of physical properties is the only constraint imposed by weak covariance.*<sup>61</sup>

Derfor kan vi konstruere eksempler der vi har mulige verdener som er fysisk uadskillelige fra den faktiske verden, men der fordelingen av mentale egenskaper er helt annerledes: I en slik

---

<sup>58</sup> Davidson (1970) s.214, (1993) s.4 og (1995) s.266.

<sup>59</sup> Kim beviser ekvivalensen i Kim (1984a) s.64. (Under visse rimelige forutsetninger.)

<sup>60</sup> Kim (1990) s.141. Kim snakker her om kovarians og ikke superveniens, fordi han i denne artikkelen ønsker å skille mellom ideene om kovarians og om avhengighet i diskusjonen. (Ibid. s.139.)

<sup>61</sup> Kim (1990) s.143.

verden kan f.eks. ingen objekter ha mentale egenskaper, eller alle objekter kan ha de samme mentale egenskapene (dette vil inkludere bl.a. gråsteiner), eller de mentale egenskaper kan være fordelt slik at mygg har menneskers mentale egenskaper og omvendt. Likefullt vil de mentale egenskapene være svakt superveniente på de fysiske ifølge vår definisjon.<sup>62</sup> Kim konkluderer med at slike scenarier utelukker at det mentale er avhengig av det fysiske.<sup>63</sup> Hvorfor hevder han dette? Vel, det er klart at et objekts mentale egenskaper ifølge svak superveniens flyter fritt i forhold til objektets fysiske egenskaper. Dette gir oss ikke avhengighet, fordi vi ikke har de *lovmessige* sammenhengene mellom fysiske og mentale egenskaper, som naturvitenskapen har gitt oss grunn til å forvente.

Men kan vi ikke hevde at vi bare er interessert i den faktiske verden, og ikke i særlige mulige verdener, og at intra-verden konsistens derfor er alt vi trenger? Problemet er at verdenene i Kims eksempel ikke behøver være særlige, for de kan dele vår verdens lover. Og i det øyeblikket naturlovene og et objekts fysiske egenskaper er på plass *må* dets mentale egenskaper være på plass. (Nomologisk nødvendighet)<sup>64</sup>

En annen form for superveniens er global superveniens. Global superveniens omhandler ikke enkelte objekter og deres egenskaper, men verdener som helhet, og kan defineres slik:

Worlds that are indiscernible in respect of subvenient properties are indiscernible in respect of supervenient properties.<sup>65</sup>

Her kreves altså ikke mer enn at i det øyeblikket en *verdens* fysiske egenskaper er fastsatt, så er dens mentale egenskaper også fastsatt. Dermed kan fysisk sett like objekter være mentalt adskillelige. Imidlertid er dette også for svakt, bl.a. fordi selv den minste forandring av baseegenskaper (Kims eksempel er "contains one more hydrogen atom") i en verden er forenelig med en hvilken som helst forandring av superveniente egenskaper: Verdenen kan f.eks. være "wholly void of mentality".<sup>66</sup>

---

<sup>62</sup> Se *ibid.* s.143, eller (1984a) s.58-60 for et mer detaljert argument.

<sup>63</sup> Kim (1990) s.143.

<sup>64</sup> Davidson hevder faktisk at svak superveniens ikke gir det uønskede scenariet med fysisk likhet men mental forskjell, siden to slike "hendelser" ville være fysisk uadskillelige men mentalt adskillelige, og dette utelukkes av svak superveniens. (Davidson (1993) s.8) Dette beror imidlertid på en klar misforståelse av Kims argument; argumentet omhandler verdener, ikke temporale hendelser i en verden. (Se Ernest Sosa (1993) s.42-43.) Kanskje er Davidson uvillig til å trekke modalitet og mulige verdener inn i diskusjonen. Men vi kan ikke forstå avhengighet uavhengig av en eller annen form for nødvendighet.

<sup>65</sup> Kim (1990) s.154.

<sup>66</sup> *Ibid.* s.155. Noen vil kanskje ha sympati med global superveniens fordi de mener en organismes historie spiller en rolle for dens mentale liv, og at dette tyder på at fysisk adskillelighet av objekter kan være forenelig med mental uadskillelighet. (Ansgar Beckermann nevner dette som et motiv (Beckermann (1992) s.98).) Spesielt er det mange som mener at mentale tilstanders *innhold* er vidt ("wide") i den forstand at det avhenger av faktorer utenfor organismen. (Se Kim (1998b) s.193-199 for en oppsummering av slike argumenter.) Derfor kan det hevdes

Diskusjonen så langt har lært oss at vår definisjon av minimal fysikalisme må inneholde en sterkere grad av nødvendighet for å tilfredsstille kravet om avhengighet.

## 2.3 Sterk superveniens

Et naturlig skritt å ta nå er å se på definisjonene av svak superveniens, og legge til at den fysiske egenskapen *nødvendigvis* er tilstrekkelig for den mentale egenskapen. Vi får da to ekvivalente definisjoner av såkalt *sterk superveniens*.<sup>67</sup> (Jeg låner navngivningen fra Kim og andre, se MPW 10.) *Modaloperatordefinisjonen*:

Mental properties *supervene* on physical properties, in that necessarily, for any mental property *M*, if anything has *M* at time *t*, there exists a physical base (or subvenient) property *P* such that it has *P* at *t*, and necessarily anything that has *P* at a time has *M* at that time. (MPW 9.)

*Uadskillelighetsdefinisjonen*:

Mental properties *supervene* on physical properties, in that necessarily any two things (in the same or different possible worlds) indiscernible in all physical properties are indiscernible in mental respects. (MPW 10.)<sup>68</sup>

Det er nå på tide å gi en nærmere fortolkning av vår superveniensrelasjon. Hva presis modaloperatordefinisjonen sier blir klarere ved litt formalisering: Hvis *x* er et objekt, *M* er en mental egenskap og **P** er basemengden av alle fysiske egenskaper, så har vi for det første som med svak superveniens at:

$$(i) \Box[(M(x) \rightarrow \exists P(P \in \mathbf{P} \wedge P(x)))]$$

(Jeg underforstår her, som ofte ellers den temporale dimensjonen i Kims definisjoner.) Så dersom *x* har *M* så har *x* nødvendigvis også én egenskap *P* i **P**. Men vi utelukker ikke at det finnes andre egenskaper i **P** som *M* kunne vært supervenient på. Dvs. vi er åpne for at det finnes en  $P^* \neq P$  slik at  $P^* \in \mathbf{P} \wedge (P^*(x) \rightarrow M(x))$ . Følgelig ville vi hatt  $M(x)$  selv om det hadde vært tilfellet at  $\neg P(x) \wedge P^*(x)$ . Dette er viktig siden vi som vi skal se må være åpne for multipl realisering, dvs. at en egenskap *M* kan ha flere mulige fysiske baseegenskaper.

---

at det mentale må være *globalt* supervenient på det fysiske. Jeg mangler både kunnskaper og plass til å behandle dette skikkelig her. Men én mulig viktig grunn for at mentalt innhold må være lokalt supervenient (dvs. supervenient på objektnivå, eller trangt/"narrow") er hvis det kun er på denne måten det kan være årsak til objektets kroppslige bevegelser. (Se f.eks. Kim (1998b) s.205-206 el. Fodor (1987) s.44.) Som vi har sett er også Davidson inneforstått med lokal superveniens.

<sup>67</sup> Forskjellen til svak superveniens er altså bare at en av kondisjonalene holder med større modal kraft. Siden definisjonene av svak superveniens er logisk ekvivalente, må dette derfor også holde for definisjonene av sterk superveniens.

<sup>68</sup> Det må være ubetydelig at han her bruker "respects" og ikke "properties." Andre steder (f.eks. Kim (1998b) s.224) formulerer han det ut fra egenskaper. Jeg har ikke funnet noe tekststed hos Kim der en eventuell nyanse mellom "respects" og "properties" ser ut til å være avgjørende.

For det andre har vi at:

(ii)  $\Box[\forall y(P(y) \rightarrow M(y))]$

Så hvis P tilfredsstillers (ii) så gjelder det altså (som vi ønsket) for ethvert objekt y at P(y) *nødvendigvis* er tilstrekkelig for M(y). Det er modaloperatoren i (ii) som skiller sterk fra svak superveniens. Som nevnt forutsetter jeg ikke en sterkere form for nødvendighet enn nomologisk nødvendighet.<sup>69</sup>

Relasjonen (i) fra M til P er dermed noe svakere enn relasjonen (ii) fra P til M. Dette er rimelig gitt at minimal fysikalisme er ment som en svak forpliktelse som må godtaes av alle som overhodet vil kalle seg fysikalister: Ellers ville definisjonen implisere bikondisjoner av typen  $M(x) \leftrightarrow P(x)$ . Det sterkeste som følger logisk nå er en potensielt uendelig lang bikondisjonal  $M(x) \leftrightarrow (P_1(x) \vee P_2(x) \vee \dots \vee P_i(x))$ : Siden disjunktene på høyre side av bikondisjonen antas å være mulige superveniensbaser for M er det klart at hver og en av dem er tilstrekkelig for, dvs. impliserer, M. Disjunksjonen av dem impliserer følgelig også M. Anta videre at M *ikke* omvendt impliserer disjunksjonen. Da må vi anta at det finnes en x slik at M(x), men  $\neg P_j(x)$  for noen  $P_j$  i disjunksjonen. M(x) er altså sann, men disjunksjonen er ikke sann. Men fordi M er en supervenient egenskap må det finnes en P slik at  $P(x) \rightarrow M(x)$ . Men en slik P er dermed en  $P_j$  i disjunksjonen, og følgelig er hele disjunksjonen likevel sann, og vi har en motsigelse. Så M må implisere disjunksjonen, og ekvivalensen holder følgelig.<sup>70</sup>

Denne ekvivalensen er rimelig gitt multippel realiserbarhet, og innebærer åpenbart ikke noen *sterk* form for reduksjonisme: Ekvivalensen gir ingen pen og ryddig sammenheng mellom det mentale og det fysiske. Jeg kommer tilbake til slike bikondisjoner i diskusjonen av emergens.

(ii) innebærer at det fysiske legger en sterk føring på det mentale. Metaforisk kan vi si at *det fysiske fikserer eller bestemmer det mentale*: Ontologisk sett betyr dette at verdens fysiske innredning nødvendigvis sikrer eksistensen av, og fastsetter verdens mentale innredning. Epistemologisk betyr det at når de fysiske setningene om et objekt har fått sin

<sup>69</sup> Man kan selvsagt med Kim se på definisjonen som generell i den forstand at man kan variere modaloperatorenes kraft i definisjonen på en måte som passer ens synspunkt (nomologisk, logisk eller metafysisk nødvendighet). Graden av nødvendighet kan også variere avhengig av hvilken type mentale egenskaper det er tale om. (MPW 10, se også Kim (1990) s.141) Det er ikke en gang påkrevet at de to modaloperatorene representerer samme type nødvendighet. (Kim (1984a) s.66) Tilsvarende kan man legge føringer på hvilke mulige verdener man aksepter i uadskillelighetsdefinisjonen. Imidlertid er vi her interessert i en minimal, og derfor formodentlig relativt svak form for nødvendighet, som nomologisk nødvendighet. Dessuten er det ikke gitt at vi bør håpe på mer enn nomologisk nødvendige sammenhenger mellom M og P: Vi er tross alt interessert i den faktiske verdens ontologi gitt dens lover. Hovedpoenget nå er at man skal kunne enes om minimal fysikalisme uavhengig av kontroverser om nødvendighetsbegrepet.

<sup>70</sup> Se f.eks. Kim (1990) s.151-152, eller også Beckermann (1992) s.97-98.

sannhetsverdi, så er det én og bare én mulig tildeling av sannhetsverdier til de psykologiske setningene om objektet. (Men superveniens i seg selv sier selvsagt ingenting om hvorvidt sannheter om det mentale kan *utledes* fra sannheter om det fysiske.) Dette er en konsekvens som blir viktig for konflikten mellom fysikk og psykologi, og det er verdt å merke seg at vi ikke ville hatt fikseringen dersom vi hadde nøyd oss med svak superveniens.

Fikseringen fremgår godt også fra uadskillelighetsdefinisjonens perspektiv på superveniens: Å si at objekter i samme eller forskjellige mulige verdener som er fysisk uadskillelige også er mentalt uadskillelige, betyr at objekter med samme fysikk må ha samme psykologi: Så snart et objekts fysiske egenskaper og verdens lover er fastsatt, er det én og bare én mulig tilordning av mentale egenskaper til objektet.

Det bør nå være relativt klart hva sterk superveniens innebærer. I MPW hevder Kim at minimal fysikalisme består i at det mentale er sterkt supervenient på det fysiske. (MPW 15.) Han forutsetter i tillegg at superveniens "inkorporerer" en avhengighets/ bestemmelsesrelasjon. (MPW 11.) Det skal altså gjøres en tilleggsantagelse om at det ligger en form for avhengighet i superveniens.<sup>71</sup> Spørsmålet vi må stille oss, er om sterk superveniens med avhengighet virkelig tilfredsstiller kravene til minimal fysikalisme.

Vi har sett at sterk superveniens tilsier at hvis x har en mental egenskap M, så har x også en fysisk egenskap P, og dermed er x i denne forstand fysisk. I.o.m avhengighetsrelasjonen antar vi i tillegg at instansieringen av M er avhengig av instansieringen av P, men ikke vice versa. Så det finnes ingen mentale egenskaper som er slik at x kunne hatt dem uten også å ha *fysiske* egenskaper. (Dvs. x's mentale egenskaper kan ikke instansieres uavhengig av x's fysiske egenskaper.)<sup>72</sup> Dermed har vi oppfylt krav (1) om monisme.<sup>73</sup>

For å oppnå asymmetrien i krav (2), må vi utelukke scenariet der det mentale og det

---

<sup>71</sup> I Kim (1998b) formulerer han derimot minimal fysikalisme som en *konjunksjon* av superveniens, et avhengighetsprinsipp og et prinsipp om at det ikke finnes ting med bare mentale egenskaper. (s.12) Dette harmonerer med kravene jeg har satt opp, men kan som vi etter hvert skal se gies en enklere formulering.

<sup>72</sup> Noen vil kanskje synes dette er urimelig av teologiske grunner: Minimal fysikalisme tilsier i så fall tilsynelatende uten videre argumentasjon at udødelige sjeler ikke finnes. Smith & Jones beroliger oss i sin lærebok med at monisme er forenelig med udødelighet som gjenoppstandelse av *dette* legeme, kun uavhengige "platoniske" sjeler utelukkes. (Smith & Jones (1986) s.29) Likeles utelukkes formodentlig bare muligheten av en transcendent gud, ikke en gud i naturen.

<sup>73</sup> Avhengighetsrelasjonen måtte brukes for å få dette resultatet. Superveniens alene er ikke nok, i hvert fall ikke hvis vi godtar at det å *ikke* ha fysiske egenskaper er en fysisk *egenskap*, eller at to ting uten fysiske egenskaper er fysisk uadskillelige. Under disse forutsetningene kan vi ha kartesianske ting med bare mentale egenskaper og samtidig opprettholde superveniens. Imidlertid påpeker Kim at slike ting (ved uadskillelighetsdefinisjonen av superveniens) må være én og samme ting, eller i det minste være helt like. (Kim (1998b) s.10) Hvilket vil være utålelig for de fleste dualister. Det er antagelig ut fra slike betraktninger Kim sier at utelukkingen av rent mentale ting ikke er "strictly implied by the supervenience principle" (ibid. s.11) og at substansdualisme ikke utelukker superveniens. (MPW 13 n.20.)

fysiske er *gjensidig* supervenient på hverandre, for i så fall måtte alle fysiske egenskaper også ha mentale baseegenskaper, hvilket åpenbart ikke gir oss asymmetri.<sup>74</sup> Men ny refleksjon over det faktum at avhengighetsrelasjonen ikke er resiprok, gjør det klart at det i det minste er mulig at enkelte objekter bare har fysiske egenskaper. Enkle empiriske undersøkelser bekrefter at så er tilfellet, og kravet er oppfylt.

Kunne vi tilfredsstilt kravet om asymmetri bare ved å postulere fravær av gjensidig superveniens, uten i tillegg å anta avhengighet? Svaret blir ”nei,” fordi vi *også* må utelukke scenariet der både P og M er superveniente på, og/eller avhengige av en tredje egenskap Q.<sup>75</sup> For følgende skjema viser at en slik Q ville gitt oss den kovariasjon av egenskaper som superveniens krever, uten at det behøver å foreligge avhengighet mellom M og P:

$$[\forall x(Q(x) \rightarrow P(x)) \wedge \forall x(Q(x) \rightarrow M(x))] \rightarrow \forall x[Q(x) \rightarrow \{M(x) \rightarrow P(x) \wedge P(x) \rightarrow M(x)\}]$$

Men fordi vi postulerer avhengighetsrelasjonen, så tilfredsstiller vi også krav (3) om avhengighet.

Altså kan sterk superveniens med en underforstått avhengighetsrelasjon tjene som definisjon av minimal fysikalisme.<sup>76</sup> Men merk at definisjonene av sterk superveniens *alene* og uten avhengighet, ikke var nok til å tilfredsstille kravet om et avhengighetsforhold mellom det mentale og det fysiske. (Jf. (2.5) nedenfor.)

## 2.4 Mereologisk superveniens

Jeg vil nå legge litt mer ontologisk struktur på verdensbildet der de mentale egenskaper er sterkt superveniente på de fysiske. Dette gjør jeg ved å introdusere den mereologiske relasjonen – dvs. relasjonen som gjelder mellom et hele og dets deler – i vår ontologi. Verden er ordnet etter denne relasjonen fra kjernepartikler, via atomer, molekyler, celler og oppover. (MPW 15-16.) Til hver grad av kompleksitet svarer et nivå (”level”) av objekter, med det Kim kaller karakteristiske egenskaper. Således er for eksempel alle objekter på det molekylære nivået sammensatt av objekter fra det atomære nivået, og har karakteristiske (spesielt kjemiske) egenskaper. Dette verdensbildet tør være relativt ukontroversielt, det er en viktig del av det naturvitenskapen har lært oss, og har som Kim sier dannet bakgrunnen for

<sup>74</sup> Kim tar opp dette scenariet i Kim (1990) s.144-145 og MPW 11.

<sup>75</sup> Kim tar opp dette scenariet i Kim (1990) s.146 og MPW 11.

<sup>76</sup> Når jeg snakker om ”superveniens,” vil jeg derfor heretter underforstå ”sterk,” hvis ikke annet er indikert.

mye metafysisk og vitenskapsfilosofisk debatt.<sup>77</sup> (MPW 16.) Notasjonsmessig lar jeg "x: [C<sub>1</sub>,...,C<sub>m</sub>; R]" bety at x består helt og holdent av ekte deler C<sub>1</sub>,...,C<sub>m</sub> organisert i relasjonen R(C<sub>1</sub>,...,C<sub>m</sub>).<sup>78</sup>

Denne ontologien gir oss ikke bare et begrepsapparat for å diskutere forholdet mellom fysiske egenskaper og spesialvitenskaper, den innebærer også en form for sterk superveniens, som kalles mereologisk superveniens. Mereologisk superveniens tilsier at et objekts makroegenskaper er superveniente på dets mikrostruktur. Kim gir en definisjon som ligner uadskillelighetsdefinisjonen av sterk superveniens:

For any *x* and *y*, belonging to level *L* (other than the lowest level), if *x* and *y* are indiscernible in relation to properties at all levels lower than *L* (or, as we may say, *x* and *y* are *microindiscernible*), then *x* and *y* are indiscernible with respect to all properties at level *L*. (MPW 17.)

Intuitivt betyr dette at en makroforskjell mellom *x* og *y* forutsetter en forskjell mellom dem på mikronivå.<sup>79</sup> Dette må vi anta holder med minst nomologisk nødvendighet, selv om dette ikke er gjort eksplisitt i definisjonen.<sup>80</sup>

Vi kan også gi en tilsvarende definisjon av mereologisk superveniens som ligner modaloperatordefinisjonen av sterk superveniens. For vi kan si at M(*x*) ved *x*: [C<sub>1</sub>,...,C<sub>m</sub>; R] er supervenient på mikrostrukturen til *x* på den måten modaloperatordefinisjonen tilsier. Men presis hvilken egenskap ved mikrostrukturen er M supervenient på? Her tror jeg vi må være forsiktige for å unngå å ta med for mye eller for lite i denne egenskapen: Det synes klart at ikke alle egenskaper ved mikrostrukturen er nødvendige for instansieringen av alle makroegenskaper.<sup>81</sup> På den andre siden må vi inkludere ikke bare atomære egenskaper ved delene, men også relasjonelle egenskaper knyttet til delenes organisering.

<sup>77</sup> Tyler Burge synes imidlertid å ha innsigelser mot at den mereologiske relasjonen kan belyse forholdet mellom det mentale og det fysiske: "It is not obvious, however, that this dependency [of the mental on the physical] need involve material constitution – the mental events being identical with or made up of physical events." (Burge (1993) s.114) På den andre siden virker det som om Burge ikke vil godta fysikalisme selv i dens *minimale* form.

<sup>78</sup> Bl.a. Beckermann bruker en slik notasjon (Beckermann (1992) s.103). Jeg ønsker ikke *a priori* å utelukke at noen objekter kan ha uendelig mange konstituenten: C<sub>1</sub>,... *ad inf.* Jeg kommer imidlertid ikke til å diskutere denne muligheten.

<sup>79</sup> Kim gir en svært presis formulering av slik mikrouadskillelighet (MPW 17.) Den bygger på at *x* og *y* er mikrouadskillelige hvis og bare hvis de har dekomposisjoner i ekte deler slik at hvis en del eller en *n*-tupple av deler i en dekomposisjon instansier *P*, så finnes det én og bare én tilsvarende del (eller *n*-tupple) i den andre dekomposisjonen som også instansierer *P*. Alternativet "n-tupple" er tatt med fordi vi må kunne åpne for superveniens på relasjonelle egenskaper. For dette se Kim (1993a) s.164.

<sup>80</sup> Men se Kim (1984a) s.66 der det faktisk gjøres eksplisitt.

<sup>81</sup> Hvis makroegenskapen f.eks. er å ha en masse på 1 kg., så er det likegyldig om objektets deler er av, la oss si vaktelegg eller bly.



Formodentlig kan det Kim kaller en mikrobasert egenskap fange opp egenskapen vi er ute etter:

*P* is a *micro-based property* just in case *P* is the property of being completely decomposable into nonoverlapping proper parts,  $a_1, a_2, \dots, a_n$  such that  $P_1(a_1), P_2(a_2), \dots, P_n(a_n)$ , and  $R(a_1, \dots, a_n)$ . (MPW 84.)

Kims eksempel er egenskapen å være et vannmolekyl som er å bestå av to hydrogenmolekyler og et oksygenmolekyl ordnet i en bestemt relasjon.<sup>82</sup> (Ibid.) Ved å fylle inn “mikrobasert egenskap” for “fysisk egenskap” i modaloperatordefinisjonen får vi definisjonen av sterk mereologisk superveniens.<sup>83</sup>

Ifølge vår definisjon blir altså kjemiske egenskaper mereologisk superveniente på mikrobaserte egenskaper ved objekter som er sammensatt av atomer. Cellulære egenskaper blir likeledes mereologisk superveniente på mikrobaserte egenskaper ved objekter som er sammensatt av molekyler osv. Vi må også følge Kim i å si at: “Mental properties, therefore, are macroproperties supervening on microproperties.” (MPW 18.)<sup>84</sup> Dette synes nødvendig, fordi vi forventer at bare nokså komplekse objekter kan utvise mentale egenskaper. Det ville være svært merkelig å tilskrive f.eks. molekyler en psykologi, selv om det selvsagt er et åpent spørsmål presist hvor grensen skal trekkes.

Til slutt vil jeg bemerke to konsekvenser av den nye superveniensrelasjonen: For det første er mereologisk superveniens en *måte* sterk superveniens kan inntreffe på. Siden det at det fysiske fikserer det mentale viste seg å være en konsekvens av sterk superveniens generelt, har vi selvsagt fikseringen også i tilfellet mereologisk superveniens.

For det andre gir mereologisk superveniens oss en formulering av superveniens og avhengighet i en pakke: Med sterk superveniens generelt måtte vi gjøre en *tilleggsantagelse* om at det skulle ligge en avhengighetsrelasjon i superveniensrelasjonen. Med mereologisk superveniens har vi derimot fått avhengigheten med på kjøpet: For det er klart at den mereologiske relasjonen som mereologisk superveniens bygger på, innebærer en form for asymmetrisk avhengighet som vi kan kalle mereologisk avhengighet. Vi vet at komplekse objekter er mereologisk avhengig av sine deler, og siden egenskaper ikke kan eksistere annet

---

<sup>82</sup> Merk at *P* som Kim påpeker (MPW 84) er en *makroegenskap*, fordi det er en egenskap som innehaes av objektet *x* og ikke av *x*'s deler. Så når jeg snakker om vitenskaper som er rettet mot mikronivået mener jeg ikke å si at disse bare studerer atomære egenskaper ved et objekts deler. Da kan også studere mikrobaserte makroegenskaper.

<sup>83</sup> Kim antyder faktisk i Kim (1984a) s.66 at den første modaloperatoren i en slik definisjon av mereologisk superveniens må representere *metafysisk* nødvendighet, mens den andre må representere nomologisk nødvendighet. Det synes da også rimelig at avhengighetsforholdet mellom makro og mikro er så sterkt at metafysisk nødvendighet er påkrevet for den første operatoren. Men intet i mitt argument avhenger av dette.

<sup>84</sup> Her kan det virke som om Kim tenker på atomære mikroegenskaper ved delene objektet er satt sammen av. Men han gjør det senere klart at selv om vi ofte snakker løselig om mereologisk superveniens, så er det han tenker på mikrobaserte egenskaper, som er *makroegenskaper*. (MPW 85-86.)

enn som instansiert av objekter (antiplatonisme) må det gjelde et tilsvarende avhengighetsforhold mellom mikrobaserte egenskaper og egenskaper som er superveniente på disse.

Mereologisk avhengighet er en form for avhengighet som Kim, i hvert fall tidligere, har antatt er metafysisk fundamental<sup>85</sup> og lovende for "mind-body supervenience".<sup>86</sup> Den gir svært klar mening til avhengigheten: Tar du bort delene, så forsvinner helheten. Likevel er mereologisk avhengighet fortsatt *minimal* i den forstand at man kan akseptere den enten man er reduktiv eller antireduktiv fysikalist. Antagelig er det mereologiske verdensbildet noe man ikke bare kan, men også *må* akseptere for å være fysikalist: For er det noe naturvitenskapen har lært oss, så er det at forklaring ut fra mikro-makro hierarkiet er svært fruktbart. Jeg konkluderer derfor med at en enhetlig og enkel definisjon av minimal fysikalisme er at mentale egenskaper er mereologisk superveniente på fysiske mikrobaserte egenskaper.<sup>87</sup>

Alle fysikalistiske veier fører derfor til mereologisk superveniens, men her skilles de avhengig av synet på forklaringer av hvorfor denne relasjonen holder. Reduksjonisten mener hun har en forklaring ( $M=P$ ), mens emergentisten, som vi skal se, i siste instans sier at det ikke er noen fullstendig forklaring på hvorfor  $M$  instansieres i den fysiske strukturen, det er et "brute fact," som bare må aksepteres. For mange antireduksjonister vil dette være en *begrunnet* oppfatning: Davidson hevder f.eks. at dersom vi mot formodning skulle komme over en streng psykofysisk lov, så ville vi ikke ha grunner til å tro at den var mer enn omtrentlig sann, fordi fysikk og psykologi har vesensforskjellige vokabularer.<sup>88</sup>

## 2.5 "Appendiks:" Forholdet mellom superveniens og avhengighet

Diskusjonen ovenfor viser at avhengighetsrelasjonen spiller en avgjørende rolle for å tilfredsstille kravene til minimal fysikalisme. Dette harmonerer med Carsten Hansens påstand om at det ikke er superveniens, men avhengighet som spiller hovedrollen i minimal fysikalisme.<sup>89</sup> Derfor er det naturlig å spørre seg: Hva er forholdet mellom avhengighet og superveniens? Kims bilde av situasjonen går ut på at det er avhengighetsrelasjonen, og ikke superveniens som forklarer forholdet mellom sinn og legeme:

---

<sup>85</sup> Kim (1993a) s.166

<sup>86</sup> Ibid. s.168

<sup>87</sup> Jeg antar at Kim vil akseptere dette som en definisjon, selv om han ikke eksplisitt gir denne definisjonen.

<sup>88</sup> Davidson (1970) s.216.

<sup>89</sup> Hansen (2000) s.459. Mer om dette følger nedenfor.

[...] mind-body supervenience itself is not an *explanatory theory*; it merely states a pattern of property covariation between the mental and the physical and points to the existence of a dependency relation between the two. Yet supervenience is silent on the nature of the dependence relation that might explain why the mental supervenes on the physical. (MPW 14.)

Dette skyldes at flere typer avhengighet “can generate the required covariation of properties and thereby qualify as a supervenience relation.” (Ibid.) Kim nevner her bl.a. “reductive dependence” og “mereological dependence” som typer avhengighet. Fordi superveniens altså ikke er “a metaphysically “deep” relation; it is only a “phenomenological” relation about patterns of property covariation” (ibid.), kan den ikke forklare kovariasjonen av egenskapene.<sup>90</sup> Forklaringen må skje ut fra det Kim kaller “deeper dependence relationships”. (Ibid.) Det som gjør sterk superveniens generelt til noe både reduksjonister og antireduksjonister kan akseptere, er selvsagt at den er forenelig med både reduktive og antireduktive avhengighetsrelasjoner. Så når Kim sier at superveniens skal inkorporere en avhengighetsrelasjon (MPW 11), så må han mene *en av flere mulige* avhengighetsrelasjoner, for som han sier; “supervenience is not a *type* of dependence relation”. (MPW 14.)

Så hvorfor ikke la avhengighetsrelasjonen alene definere minimal fysikalisme? Som Hansen viser, er det avhengighet som forklarer og impliserer superveniens, og det er angående hva slags avhengighet som gjelder at forskjellige fysikalister er uenige. Derfor hevder han at superveniens er overflødig i en definisjon av minimal fysikalisme,<sup>91</sup> og sier: “On reflection, mind-body dependence seems not only necessary, but sufficient for physicalism.”<sup>92</sup>

Det er likevel grunner til å følge Kim i å definere minimal fysikalisme ut fra superveniens, og ikke bare ut fra avhengighet. For det første er superveniens et svært nyttig og presist konseptuelt verktøy, hva gjelder egenskapsinstansieringer. Det bringer klart og umiddelbart frem forhold mellom objekters super- og subveniente egenskaper. For det andre har vi sett at Kim har valgt en bestemt form for avhengighet, nemlig mereologisk avhengighet og dermed endt opp med mereologisk superveniens. I mereologisk superveniens er derfor (den mereologiske) avhengigheten *bygget inn i* superveniensbegrepet som en naturlig del. Derfor vil definisjonen av minimal fysikalisme inneholde kun en komponent (mereologisk superveniens), i stedet for to (supervenience og avhengighet). Det er altså rimelig å ikke erstatte snakk om superveniens med snakk om avhengighet. Men moralen fra Hansens

---

<sup>90</sup> Han gjør et tilsvarende poeng med samme “dybdemetafor” i Kim (1990) s.147.

<sup>91</sup> Hansen (2000) s.460.

<sup>92</sup> Ibid. Hansen aksepterer at superveniens kan fungere som en “[...] *working* definition – for the purpose of setting different positions in relation to each other, and for framing further questions [...]” (Ibid. s.461.) Dette fordi fysikalisme er forpliktet til superveniens (ibid. s.460), men superveniens fanger likevel ikke kjerneideen i minimal fysikalisme. (Ibid. s.461)

artikkel og diskusjonen ovenfor er at det er avhengighetsrelasjonen i superveniens som gjør den argumentative jobben, superveniens alene er bare et konseptuelt verktøy.<sup>93</sup>

## 2.6 Oppsummering

Jeg har argumentert for at mereologisk superveniens definerer minimal fysikalisme. Gjennom dette har jeg beskrevet en ontologi for en verden som er organisert nivåvis etter del-hele relasjonen, og med den viktige konsekvens at det mentale – *qua* høyere nivå – i en bestemt forstand *fikses* av det lavere fysiske nivå. Når vi i tillegg antar en bestemt form for realisme, så vil denne konsekvensen skape problemer for psykologiens autonome status.

## 3 Intuisjon om realisme

### 3.1 Innledning

Mitt utgangspunkt her er en intuisjon om realisme: Dersom psykologiens forklaringer skal ha sannhet og gyldighet, så må psykologien forstås som en vitenskap om trekk ved verden. Fra dette utgangspunktet formulerer jeg en form for egenskapsrealisme som tilsier at psykologiske forklaringer har realitet hvis mentale *egenskaper* har kausale krefter. Et lignende syn på psykologiske forklaringer finner vi også hos Kim. Det vil imidlertid vise seg at Davidsons analyse av hendelser og kausalrelasjoner mellom hendelser, er i konflikt med dette synet.

Jeg sier ”intuisjon” om realisme fordi realismen danner et utgangspunkt for diskusjonen, og fordi jeg ikke gir noen veldig omfattende begrunnelse for den. Forutsetningen vil sikkert virke rimelig for mange: En eller annen form for realisme kan være et tungtveiende metafilosofisk hensyn, som det må svært viktige argumenter til før man gir slipp på. Men motstandere av realismen for psykologi kan følge argumentasjonen for å se hvor antagelsen fører hen, eventuelt med tanke på et *reductio ad absurdum*: For det vil bli klart at problemet jeg motiverer for psykologiens status ikke oppstår uten antagelsen om realisme, og dersom

---

<sup>93</sup> Selv om Kim vil si seg enig i dette, er han tildels uklar når det gjelder forholdet mellom superveniens og avhengighet: “What needs to be added to property covariance to get dependence or determination, or whether dependence/determination must be taken as an independent primitive [as *Kim* actually does in his (1998b) p. 11], are difficult questions that probably have no clear answers.” (MPW 11.) Hansen finner Kims skepsis m.h.t muligheten av et klart svar på dette spørsmålet merkelig, fordi det tyder på at vi ikke *kan* få en definisjon av minimal fysikalisme med superveniens som en ”primitive component”. (Hansen (2000) s.460) Merkelig er det også, men jeg tror at vi pga. velvillighetshensyn bør forstå sitatet som at superveniensavhengighet ikke er en *type* avhengighet (se ovenfor og Kim (1993a) s.166), og at det *derfor* ikke finnes noe klart svar på hva mer en kovariasjon som kreves for å gi avhengighet til superveniens *generelt*. I *enkelttilfellet* mereologisk superveniens er mereologisk avhengighet som vi har sett et klart svar. Dette stemmer også med at Kim har gitt uttrykk for et metodologisk motiv om å ha mange generelle superveniensbegreper som kan tilpasses til forskjellige anvendelsesområder. (Kim (1990) s.155)

problemet ikke kan gies en tilfredsstillende løsning, så kan dette kanskje for noen være en god nok grunn til å forkaste realismen.

### 3.2 Egenskaper og kausale krefter

Vi har sett (1.2-1.3) at psykologi og andre spesialvitenskaper kan betraktes som begrepslige praksiser kjennetegnet ved (i) karakteristiske begreper, og (ii) prinsipper/mulighetsbetingelser for anvendelsen av disse. I psykologiens tilfelle er eksempler på begreper ”tror at,” ”ønsker at” osv. Disse er knyttet til påstandsholdninger og forutsetter en bakgrunn av rasjonalitet.

Jeg tar utgangspunkt i to krav til psykologiens status: (1) For at psykologi skal kunne ha noen grad av sannhet og gyldighet må den forstås som en teori *om* en verden som er grunnleggende *fysisk* i den forstand at alle objekter har fysiske egenskaper. Psykologi omhandler altså objekter som også omhandles av fysikk. Enkelte vil selvsagt hevde at psykologien bare er en nyttig modell som kanskje ikke refererer til faktiske trekk ved verden. Men denne formen for instrumentalisme vil jeg altså unngå her. (2) For at krav (1) skal tilfredsstilles må psykologiens begreper, prinsipper og mulighetsbetingelser i en eller annen forstand kunne *føres tilbake til trekk ved den fysiske verden*. Det må være noe i verden som gjør psykologiske sannheter sanne. Vi må altså finne en metafysisk relasjon mellom psykologien – forstått som en begrepslig praksis – og verden. Gitt den prinsipielle forskjellen mellom fysikk og psykologi er det klart at dette er et strengt krav. Men på den annen side behøver ikke trekket ved den fysiske verden være et essensielt fysisk trekk, alt vi krever er at tilbakeføringen er forenelig med supervenienstesen i minimal fysikalisme.

Nå er dette med tilbakeføring foreløpig mildest talt generelt og uklart. Men det finnes i hvert fall én rimelig klar og sentral måte dette kravet til psykologien kan oppfylles på: *Hvis begrepene i dens forklaringer refererer til egenskaper som har kausale krefter er dette en tilbakeføring til trekk ved verden*. Hvorfor vil dette gi realitet til psykologi? For det første er det å være reell tett, om ikke også essensielt, knyttet til det å ha kausale krefter.<sup>94</sup> Under forutsetning av en kausal persepsjonsmodell, og at vi kun får kjennskap til noe gjennom dets (kanskje indirekte) effekter på våre sanseorganer, så følger det at vi ikke kan ha perseptuell viten om ting som mangler kausale krefter.<sup>95</sup> For det andre spiller kausale forklaringer (f.eks. av handlinger som forårsaket av ønsker og oppfatninger) åpenbart en uhyre viktig rolle for

---

<sup>94</sup> Kim er blant de som ser på besittelse av kausale krefter som et plausibelt kriterium for realitet. (MPW 119.) Se også f.eks. Kim (1989b) s.279.

<sup>95</sup> Dette poenget er inspirert av Kims kommentar om at det er usikkert om det gir mening å snakke om epifenomener i streng forstand (dvs. om fenomener som er effekter men ikke årsaker), fordi ”it is doubtful how such events could be known to exist.” (Kim (1984b) s.94)

psykologien.<sup>96</sup>

Ville en svakere form for tilbakeføring kunne tilfredsstillende krav (1) om at psykologi må være *om* den fysiske verden? Kanskje. Men en tilbakeføring som ikke gir det mentale en konkret plass i verden ved å postulere eksistensen av mentale egenskaper står i fare for å overlate for mye til fysikken, og dermed true psykologiens *autonomi*. Konkret skal vi se (6.4) at f.eks. funksjonalisme antagelig ikke er forenelig med psykologi som en strengt autonom vitenskap, og at en sterkere form for egenskapsdualisme må til for å sikre psykologiens autonomi.

Hva jeg forplikter meg til med hensyn til forholdet mellom egenskaper og kausale krefter kan sees ved et eksempel: Lille Per slår pappa i hodet med balltreet, *og slaget med balltreet forårsaker en hodepine*. Her betraktes hendelsene som egenskapsinstansieringer: Det er *fordi* (alternativt i kraft av at el.l.) den første hendelsen består i visse egenskapsinstansieringer (formodentlig bl.a. at balltreet instansierer egenskapen hardhet) at den andre hendelsen består i en eller flere egenskapsinstansieringer (pappa instansierer egenskapen å ha vondt i hodet). Instansieringer forårsaker hverandre.

Vi kan forklare hodepinen ved å bruke begrepet ”hardt,” og forklaringen har realitet, fordi begrepet kan føres tilbake til en (i dette tilfellet kausalt relevant) egenskap i verden.

Dette kan synes både enkelt og trivielt. Grunnideen i eksempelet støttes av refleksjon over kontrafaktiske situasjoner: Hvis ikke balltreet hadde vært så hardt, så hadde ikke pappa fått vondt i hodet. Det må bemerkes at jeg ikke forutsetter noen metafysisk tung teori om egenskaper: Intet i mitt argument avhenger av at egenskaper er *noe mer enn* et sett med kausale krefter. Følgelig er det heller ikke slik at vi kan forklare *hvorfor* et objekt har visse kausale krefter ved å henvise til at det har bestemte egenskaper. En annen konsekvens av dette er at egenskaper er identiske hvis og bare hvis de har samme kausale krefter.<sup>97</sup>

Siden jeg her analyserer hendelser som egenskapsinstansieringer ved et tidspunkt, får vi også følgende identitetskriterium for *hendelser*:

---

<sup>96</sup> Etter Davidsons artikkel ”Actions, Reasons, and Causes” (1963) er det igjen blitt filosofisk respektabelt å betrakte (i det minste *noen*) psykologiske forklaringer som kausalforklaringer. Men den sterkere tesen at *alle* psykologiske forklaringer i siste instans er kausale behøves ikke. La oss bare si at psykologiens realitet ville være svært alvorlig skadet, hvis mentale egenskaper ikke kan være årsaker.

<sup>97</sup> Strengt tatt bør jeg derfor ikke si at egenskaper ”har” kausale krefter, men at de ”er” dem. Når jeg av bekvemmelighetshensyn likevel noen ganger skriver på denne måten må det huskes at jeg forutsetter et svakt begrep om ”egenskap.”

Hendelsen H der objektet x instansierer egenskapen P ved tiden t er identisk med hendelsen H\* der x\* instansierer P\* ved t\* hvis og bare hvis  $x=x^*$  &  $P=P^*$  &  $t=t^*$ .<sup>98</sup>

Det synet på hendelser, egenskaper og kausalsammenhenger som jeg har skissert, står på mange måter i konflikt med Davidsons syn. Et symptom på denne konflikten er at Davidson som kjent tidligere har forsøkt å forfekte et ganske annet identitetskriterium for hendelser:

[...] events are identical if and only if they have exactly the same causes and effects.<sup>99</sup>

Jeg vil nå forsøke å motivere at det å betrakte hendelser som egenskapsinstansieringer er rimelig gitt visse trekk ved våre forklaringspraksiser. Likeledes vil jeg få frem at det er vanskelig å se hvordan Davidsons syn på hendelser er forenelig med disse praksisene.

Det er viktig å avgrense analysen av hendelser som egenskapsinstansieringer fra den absurde tesen om at beskrivelser av hendelser forandrer deres kausale sammenhenger. Davidson har utvilsomt rett når han i artikkelen "Thinking Causes" (1993) sier at kausalitet er en ekstensjonal relasjon, dvs. en relasjon som holder uavhengig av hva som måtte skje på et intensjonalt nivå eller beskrivelsesnivå. Hvordan vi beskriver årsak og virkning påvirker selvsagt ikke kausalrelasjonen:

[...] events are non-abstract particulars, and [...] causal relations are extensional relations between such events. [...] But there is then no room for a concept of 'cause as' which would make causality a relation among three or four entities rather than between two. On the view of events and causality assumed here, it makes no more sense to say event *c* caused event *e* as instantiating law *l*<sup>100</sup> than it makes to say *a* weighs less than *b* as belonging to sort *s*. If causality is a relation between events, it holds between them no matter how they are described.<sup>101</sup>

Imidlertid tar han feil når han slutter fra beskrivelsers kausale insignifikans til at det ikke gir mening å si at noe forårsakes i kraft av eller pga. en egenskap. Eksempelet om lille Per er et eksempel på at dette er mulig. Som Brian P. McLaughlin sier, viser heller ikke Davidsons eksempel på en ekstensjonal relasjon – relasjonen "veier mindre enn" – umuligheten av dette: For riktignok vil *a* veie mindre enn *b* uavhengig av hvilken "sort" vi beskriver *a* som (f.eks. "a er fjærvektsbokser" eller "a er født i Kautokeino"), men relasjonen kan likevel holde *fordi* *a* og *b* har visse egenskaper, nemlig vekter *v* og *w* slik at *v* er mindre enn *w*.<sup>102</sup> En forklaring av

---

<sup>98</sup> Jf. Kim (1976) s.35.

<sup>99</sup> Davidson (1969) s.179.

<sup>100</sup> Davidson diskuterer hvorvidt noe kan sies å forårsake noe forsåvidt som, eller fordi det instansierer en lov.

<sup>101</sup> Davidson (1993) s.6.

<sup>102</sup> McLaughlin (1993) s.31-32. McLaughlin formulerer det som at relasjonen holder pga. et "fact." (Ibid. s.32) Dersom man har innsigelser mot å si at relasjonen holder pga. egenskaper *v* og *w* ved henholdsvis *a* og *b*, så må man gjerne snakke om en relasjonell egenskap *R(a,b)*.

relasjonen må derfor i siste instans vise til egenskapene v og w. (Jeg sier ”i siste instans” fordi forklaringer som ikke eksplisitt viser til v og w også kan ha realitet: Jeg kan f.eks. forklare at a veier mindre enn b ved å si at ”a oppnådde bedre resultater på slankekurset enn b gjorde.” Men jeg håper man vil være enig med meg i at denne forklaringen i siste instans holder pga. at a instansierer vekten v og b instansierer vekten w.)

Synet på kausalforklaring jeg forfekter her, avhenger kun av at predikater i våre beskrivelser noen ganger (forhåpentligvis nokså ofte) refererer til egenskaper forstått som kausalpotente trekk ved verden, og at disse egenskapene er kausalt virksomme i den gitte situasjonen. Dersom dette er tilfellet, har vi en gyldig kausal forklaring. I mitt eksempel refererer beskrivelsen ”å være hard” til en slik egenskap, mens ”å være et balltre som pappa har laget til Per” eller ”å være et brunt balltre” (formodentlig) ikke gjør det.<sup>103</sup> Selv om alle beskrivelsene er sanne om det samme objektet er bare den første egnet til å inngå i en kausalforklaring.

Men Davidson betrakter som kjent hendelser som primitive ontologiske entiteter på linje med objekter.<sup>104</sup> Han vil ikke som Kim og analysere hendelsers indre struktur ved si at hendelser *er* egenskapsinstansieringer av et objekt ved et tidspunkt.<sup>105</sup> Det er formodentlig dette som i sitatet ovenfor får ham til å fornekte ”a concept of ‘cause as’” dvs. benekte at det gir mening å si at en hendelse forårsaker noe fordi den består i en instansiering av en kausalt virksom egenskap slik jeg forutsetter. Som vi så er kausalrelasjonen for ham en ekstensjonal relasjon mellom uanalyserte hendelser betraktet som individer.

Imidlertid tror jeg ikke Davidson vil gå så langt som å si at kausalrelasjoner mellom hendelser er et helt uforklarlig faktum. Han hevder nemlig også i ”Thinking Causes” at man fra *fysiske* hendelsesbeskrivelser sammen med lover kan dedusere den kausale relasjonen, mens dette ikke er tilfellet for f.eks. mentale beskrivelser.<sup>106</sup> Det synes derfor naturlig å spørre seg hva som for Davidson er forholdet mellom (i) kausalforklaringer som tillater deduksjon av kausalrelasjonen, og (ii) kausalforklaringer som ikke tillater slik deduksjon.

Et eksempel fra en tidligere artikkel kan belyse dette: Davidson sier at hendelsen ”the

---

<sup>103</sup> Selv i dette enkle tilfellet spiller flere egenskaper en kausal rolle. Eksempler vil være hodets relative mykhet, Pers relative styrke etc. Det er derfor viktig å presisere at denne realismen om forklaring ikke innebærer noe krav om *fullstendighet* (svært få, hvis noen forklaringer er fullstendige, de nevner ikke alle virksomme faktorer eller egenskaper) eller om *infallibilitet* (kanskje var pappa i bakrus og kanskje slo ikke Per hardt; våre forklaringer er ofte feilaktige). Skjemaer, definisjoner o.l. i analytisk filosofi der man f.eks. ganske enkelt viser til en mental egenskap M og en fysisk egenskap P kan være villedende enkle: De fysiske mikrobaserte egenskapene kan ofte forventes å være ekstremt komplekse.

<sup>104</sup> Se f.eks. Davidson (1969) s.175.

<sup>105</sup> Se f.eks. Davidson (1969) s.170, Kim (1976) s.35

<sup>106</sup> Davidson (1993) s.6



striking of the match” forårsaker hendelsen ”the lighting of the match.”<sup>107</sup> En beskrivelse som nevner nødvendige kausale faktorer som tørrhet, nærvær av oksygen etc. kan prinsipielt inngå i en deduksjon av virkningen, mens en beskrivelse som bare nevner ”the striking,” selvsagt ikke er tilstrekkelig for en slik deduksjon. Like fullt insisterer Davidson på at hendelsen ”the striking of the match” er årsaken og ikke bare er *en del av* årsaken, fordi hendelsen ”the striking of the match” er identisk med en hendelse der alle kausalt relevante faktorer er på plass: ”In fact, it is not *events* that are necessary or sufficient as causes, but events as *described* in one way or another.”<sup>108</sup> Dermed er det bare beskrivelsen som kan være utilstrekkelig, ikke hendelsen. Hendelsen er det den er.<sup>109</sup> Så når Davidson skal innfri sitt løfte om at psykologiske forklaringer kan være kausalforklaringer, blir det klart at disse forklaringene ikke er bedre stilt enn en forklaring av ”the lighting of the match” som bare nevner ”the striking of the match.” Mentale hendelser er altså årsaker bare forsåvidt som de er identiske med hendelser som har fysiske beskrivelser. Disse fysiske beskrivelsene kan i sin tur prinsipielt inngå i en deduksjon av virkningen.

Grunnen til at dette synes utålelig, er ikke at forklaringer må være fullstendige eller tillate deduksjon for å ha praktisk verdi. Problemet er snarere følgende: Det at en forklaring er komplett eller tillater oss å dedusere den kausale relasjonen tyder på at *den speiler verden mer nøyaktig* enn andre forklaringer. En slik forklaring virker så godt *fordi* den representerer de kausalt relevante trekkene ved situasjonen. At den gjentatte ganger skulle virke så godt *uten* å avbilde trekkene ville være en tilfældighet på kosmisk nivå. (Jeg ser nå bort fra velkjente skeptiske argumenter.) I speilbildemetaforen ligger det imidlertid en grunnleggende intuisjon om en form for realisme som Davidson neppe vil være med på.

Slike betraktninger kan gi opphav til en bekymring om at det er fysiske forklaringer som avbilder de virksomme egenskapene, og at psykologiske forklaringer blir overflødige.

Men også uavhengig av disse resonnementene ser det ut til at common sense og forklaringsmessig praksis i det minste her synes mer på linje med en ontologi av hendelser som egenskapsinstansieringer enn med en ontologi der hendelser er primitive og uanalyserte.<sup>110</sup>

Imidlertid går Davidson til en viss grad med på å diskutere hvordan egenskaper –

---

<sup>107</sup> Davidson (1969) s.172.

<sup>108</sup> Ibid.

<sup>109</sup> Jf. også f.eks. Davidson (1967) s.155-156.

<sup>110</sup> Dette blir stående som en påstand eller intuisjon jeg ikke begrunner ytterligere. Debatten omkring hendelser vs. egenskapsinstansieringer er omfattende og teknisk, fordi hvilken hendelsesontologi vi bør ha er avhengig av hvordan setninger som omhandler hendelser skal formaliseres.

forstått som beskrivelser<sup>111</sup> – gjør en kausal forskjell for hendelser. I ”Thinking Causes” ønsker han nemlig å tilbakevise at hans egen teori impliserer at ”the mental properties of an event make no difference to its causal relations”.<sup>112</sup> Kim kommenterer:

In fact, Davidson himself acknowledges in the end that it makes sense to discuss the causal relevance of properties; for, after all, he offers an account of it, based on supervenience and non-strict laws.<sup>113</sup>

Her er Davidson:

For supervenience as I have defined it does, as we have seen, imply that if two events differ in their psychological properties, they differ in their physical properties (which we assume to be causally efficacious). If supervenience holds, psychological properties make a difference to the causal relations of an event, for they matter to the physical properties, and the physical properties matter to the causal relations.<sup>114</sup>

*Hvilke beskrivelser som er sanne om hendelser* gjør altså en forskjell for kausalrelasjonene: Mentale egenskaper gjør en kausal forskjell for såvidt som de er svakt superveniente på fysiske egenskaper som er ”causally efficacious,” formodentlig i den forstand at de prinsipielt kan inngå i en deduksjon av kausalrelasjonen. Men det å være svakt supervenient på fysiske egenskaper, er kun å gjøre en kausal ”forskjell” i en *meget* svak forstand av ordet.<sup>115</sup> Davidson blir dermed likevel sårbar for eksklusjonsbaserte argumenter der det vises nettopp at det er fysiske og ikke mentale egenskaper som gjør den virkelige kausale jobben. Og som svar på argumenter om at ”forskjellen” blir for svak, synes Davidson ganske enkelt å insistere på at en hendelse med en supervenient egenskap og en (ellers lik) hendelse uten denne egenskapen, rett og slett ikke er samme hendelse.<sup>116</sup> Hvis dette er det sterkeste han kan gå med på, så er det uklart om han burde snakke om egenskapers kausale relevans i det hele tatt. Å hevde at kausale relasjoner er ”brute facts” ved hendelser ville kanskje vært en mer stabil posisjon.<sup>117</sup>

---

<sup>111</sup> Davidson (1993) s.4.n.3.

<sup>112</sup> Ibid. s.13.

<sup>113</sup> Kim (1993a) s.22

<sup>114</sup> Davidson (1993) s.14.

<sup>115</sup> Dette utdypes i Sosa (1993) s.43-46.

<sup>116</sup> Davidson (1993) s.17.

<sup>117</sup> Kim nevner dette som en mulig posisjon hva gjelder hendelsers kausale krefter. (Kim (1993a) s.22.)

### 3.3 Prinsippet om det mentales realitet

Jeg har skissert et syn på hendelser, egenskaper og kausal relevans som kan sies å utgjøre en form for egenskapsrealisme. Egenskapsrealismen kan oppsummeres slik: Anta generelt at "M" er et (potensielt komplekst) psykologisk uttrykk som skal forklare at et objekt  $x$  instansierer en (mental eller fysisk, og potensielt kompleks) egenskap  $F$ . Vi kan da formulere følgende prinsipp om realitet for psykologiske forklaringer, som jeg kaller *prinsippet om det mentales realitet*:

Dersom  $F(x)$  skal kausalforklares av "M," så må "M" referere til en mental egenskap  $M$  som instansieres av et objekt  $y$  (som muligens er lik  $x$ ), og  $M(y)$  må forårsake  $F(x)$ .

Siden Kim står sentralt i denne oppgaven, er det verdt å merke seg at diskusjonen i dette kapittelet, som hittil har vært nokså uavhengig av Kim, har brakt oss til sentrale forutsetninger i hans filosofiske posisjon. Det vil nemlig bli klart at den argumentative kjernen i MPW omhandler nettopp mentale egenskapers kausale krefter. (7.2) Det er også klart at Kim er realist når det gjelder kausalforklaringer og egenskaper. Følgende sitat korresponderer i flere viktige henseender med prinsippet jeg har formulert:

For the only way in which I believe that we can understand the idea of causal explanation presupposes the idea that the event invoked in a causal explanation is in reality a cause of the phenomenon to be explained. That is, if  $c$  (or a description or representation of  $c$ ) causally explains  $e$ ,  $c$  must *be* a cause of  $e$ . If my desire for a drink of water causally explains my body's movement toward the kitchen, the desire must really be a cause of the bodily movement. I take this to be an untendentious and uncontroversial point. (MPW 64.)

Kim innrømmer at "forklaring" er et løst begrep men insisterer likevel på dette punkt; at det som skal *kausalforklare* faktisk må være årsaken.<sup>118</sup> (MPW 76.) Sitatet ovenfor er formulert ut fra hendelser, men kan for Kim lett oversettes til snakk om egenskapsinstansieringer.

Følgende konklusjoner kan trekkes: Psykologiens antatte realitet avhenger i høy grad av at dens egenskaper kan være årsaker. Spørsmålet om psykologiens realitet (og som vi skal se; autonomi) er dermed nært knyttet til spørsmålet om mentale egenskapers kausale krefter, og dermed til den videre filosofiske problematikken omkring "mental causation." Videre vet vi at metafysikken omkring disse egenskapene må være forenelig med det vi har sagt om minimal fysikalisme, dvs. med sterk mereologisk superveniens.

---

<sup>118</sup> Et annet sted i Kims verker finner vi den mer generelle påstanden om at relasjoner i verden må underligge våre teorier og forklaringer. (Kim (1984a) s.53).

## 4 Problemet for psykologiens status

### 4.1 Innledning

Jeg er nå kommet til et punkt hvor problemet jeg motiverer for psykologiens og det mentales status kan formuleres. Det vil også vise seg at emergentisme ser ut til å kunne løse dette problemet.

### 4.2 Konflikten mellom fiksering og autonomi

Problemet oppstår fra de tre forutsetningene jeg har beskrevet og delvis begrunnet, nemlig (1) *at psykologi og fysikk er vesensforskjellige*, (2) *minimal fysikalisme* og (3) *egenskapsrealisme for psykologi*.

Vi fant ved (1) at psykologi og fysikk er vesensforskjellige disipliner, eller begrepslige praksiser: Prinsippene for anvendelsen av deres respektive begreper er motstridende, og derfor kan ikke psykologi reduseres definatorisk til fysikk. Så psykologi er autonom i forhold til fysikk, i den minimale forstand at den opererer uavhengig av fysikken. Denne uavhengigheten ekspliserte jeg i (1.3).<sup>119</sup>

Imidlertid innebærer (2) minimal fysikalisme en forpliktelse til sterk superveniens: Dette betyr som vi har sett at det mentale blir *fiksert* at det fysiske: Ontologisk betyr dette at mentale egenskaper avhenger av fysiske egenskaper. De (superveniente) mentale egenskapene instansieres nødvendigvis når passende (subveniente) fysiske egenskaper instansieres. Epistemologisk betyr dette at når de fysiske sannhetene om et objekt er på plass, så er også de mentale sannhetene om objektet på plass. Med sannheter mener jeg her sanne setninger: Når setningene om et objekts fysiske egenskaper har fått sin sannhetsverdi, så er det én og bare én mulig tildeling av sannhetsverdier til setningene om dets mentale egenskaper. Men det er foreløpig et åpent spørsmål om det finnes systematiske sammenhenger mellom de fysiske og de psykologiske setningene. Minimal fysikalisme forutsetter ikke at man kan slutte fra fysiske beskrivelser av et objekt, til psykologiske beskrivelser av det.

Alene står ikke (1) og (2) i noen konflikt, men vi antok i tillegg (3) at psykologi måtte tilbakeføres til trekk ved verden hvis den skulle ha realitet: Psykologiske begreper må referere til kausalt virksomme egenskaper i verden. Denne antagelsen genererer en spenning mellom (1) og (2).

---

<sup>119</sup> La meg for sikkerhets skyld igjen understreke at dette ikke innebærer at fysiske begreper ikke kan inngå i psykologiske forklaringer, og at "motstridende" ikke innebærer logisk motsigelse.

Før vi går inn på hvilke følger våre tre forutsetninger får for psykologiens status som en autonom vitenskap, må vi imidlertid se litt nærmere på hva vi bør forstå med begrepet autonomi. På grunnlag av egenskapsrealismen og karakteristikken av vitenskaper som begrepslige praksiser synes det følgende som et rimelig, om enn sterkt krav: En autonom spesialvitenskap (f.eks. psykologi) må si *noe mer enn, eller forskjellig fra det fysikk sier*. Innenfor det egenskapsrealistiske perspektivet betyr ”mer enn” at vitenskapen må referere til ikke-fysiske egenskaper. Dermed er den historien spesialvitenskapen forteller oss også ”*forskjellig fra*” fysikkens historie: For det at egenskapene er ikke-fysiske, innebærer at de må være *egenskaper av en annen art enn* de fysiske egenskapene. Denne forskjellen ligger i at prinsippene for beskrivelse av de spesialvitenskapelige egenskapene, er forskjellige fra fysiske prinsipper, slik vi har sett i forbindelse med psykologi og fysikk. Slik forstått vil en autonom vitenskap altså fortelle oss om ikke-fysiske egenskaper, som den beskriver på en ikke-fysisk måte, dvs. ut fra ikke-fysiske prinsipper. Dermed vil denne vitenskapen ha et eget og autonomt sett med egenskaper i sitt diskursunivers, og den vil gi oss viten som fysikken ikke gir oss.<sup>120</sup>

Hvis en vitenskaps karakteristiske egenskaper derimot kan gies en *fullstendig* fysisk forklaring, så er det klart at denne vitenskapen ikke forteller oss noe nytt om verden, og følgelig er den heller ikke autonom. Et annet kriterium for vitenskapers autonomi blir derfor at *en vitenskap er autonom hvis den er utenfor fysikkens forklaringsmessige rekkevidde*.

For å vende tilbake til spørsmålet om psykologiens status så krever (3) (realismen) at psykologien må gies en metafysisk forankring i den *fysiske* verden. Dette innebærer at mentale egenskaper må være egenskaper ved fysiske objekter, dvs. ved objekter som også har fysiske egenskaper. Men hva er den metafysiske relasjonen mellom objektenes mentale og fysiske egenskaper?

Dette spørsmålet gir – som vi så i innledningen – opphav til et slags ”dilemma.” ”Dilemmaet” er ikke formelt, eller logisk sett et dilemma, men uttrykker hva som kan skje dersom relasjonen mellom mentale og fysiske egenskaper blir for svak, eller for sterk: For det første kan det tenkes at relasjonen blir for svak. Vi lykkes rett og slett ikke i å forankre de mentale egenskapene i den fysiske verden på en slik måte at de gjør en reell kausal forskjell. En måte dette kunne skje på er hvis psykologi rett og slett viser seg å være *for* vesensforskjellig fra fysikk eller moden nevrovitenskap til å kunne være om den fysiske/nevrofysiologiske verden i det hele tatt. Hvis vi da ikke vil revurdere våre fysikalistiske

---

<sup>120</sup> Jeg bruker Paul Oppenheim og Hilary Putnams uttrykk ”diskursunivers” om det område av virkeligheten (de objekter og egenskaper) en vitenskap omhandler. (Oppenheim & Putnam (1958) s.6))

forpliktelser, så må psykologiens forskjell fra fysikken tale i psykologiens disfavør: Enten er den et praktisk – ja, kanskje til og med uunværlig – forklaringsverktøy som ikke (eller antageligvis ikke) refererer til trekk ved verden. (Instrumentalisme.) Eller så finnes det ikke noe mentalt, psykologien er usann, og må elimineres til fordel for fysikk. (Eliminativisme.) En tredje mulighet er selvsagt at vi finner en relasjon som er for svak til at det mentale kan gjøre en kausal forskjell. (Epifenomenalisme.) Hvis det imidlertid forutsettes at det å være reell er å gjøre en kausal forskjell (3.2), så er epifenomenalisme kanskje ikke så forskjellig fra eliminativisme. (Jf. MPW 119.) Ingen av disse alternativene gir psykologi den form for autonomi vi er ute etter.

Imidlertid krever (2) at hvis det finnes mentale egenskaper, så må relasjonen mellom dem og fysiske egenskaper være minst like sterk som sterk superveniens. Dette innebærer som vi har sett at det mentale fikseres av det fysiske: Fysiske egenskaper er nødvendigvis tilstrekkelige for mentale egenskaper. Trusselen med dette er at den metafysiske relasjonen mellom det mentale og det fysiske blir så tett at det mentale kommer innenfor fysikkens forklaringsmessige rekkevidde. (Reduksjonisme.) Dermed er psykologien heller ikke her autonom. Det er dette siste hornet av dilemmaet jeg behandler her.

*Konflikten står dermed mellom fiksering og autonomi: Hvordan kan den psykologiske måten å snakke på være både fiksert av, og autonom i forhold til den fysiske? Hvilken plass i verden kan psykologiske egenskaper ha, hvis de skal være fiksert av fysiske egenskaper, men samtidig være forskjellige fra dem, og ha egne kausale krefter?*

Vi må altså finne en stabil (antireduktiv) posisjon mellom instrumentalisme, eliminativisme og epifenomenalisme på den ene siden og reduksjonisme på den andre siden.<sup>121</sup> Kims kommentar om at vår moderne versjon av "the mind body-problem" er "that of finding a place for the mind in a world that is fundamentally physical." (MPW 2) viser seg å være en presis diagnose av den balansegang som kreves for å løse spenningen mellom fiksering og autonomi.

#### **4.3 Emergentisme kan tilsynelatende løse konflikten**

Hvis vi ser på den ontologiske strukturen disse spørsmålene er blitt stilt innenfor, kan vi se mer presist hvordan konflikten mellom fiksering og autonomi oppstår. Men vi kan også se

---

<sup>121</sup> Antirealisme er vel å merke en løsning, for uten (3) oppstår ikke problemet. (I en viss forstand kan instrumentalisme være en form for antirealisme. Imidlertid kan man komme frem til instrumentalismen fra et realistisk utgangspunkt, der man først prøver å finne en plass for det mentale i verden, men ikke lykkes. Hvis man derimot er antirealist og instrumentalist i utgangspunktet, vil man ikke en gang begynne på prosjektet med å lete etter en plass for mentale egenskaper i den fysiske verden.)

kimen til en løsning av den. Diskusjonen om minimal fysikalisme startet fra sterk superveniens, men vi la på mer ontologisk struktur til mereologisk superveniens. Påstanden om fiksering blir innenfor den mereologiske modellen av en lagdelt verden til påstanden om at det fysiske mikronivået fikserer makronivåene. Likeledes blir spørsmålet om autonomi til spørsmålet om det kan finnes disipliner som har makroobjekter og deres egenskaper som sitt diskursunivers, og som samtidig er autonome i forhold til disiplinene som har mikronivå som sitt diskursunivers.<sup>122</sup>

Innenfor det mereologiske verdensbildet kan vi altså se tydeligere hvordan fikseringen truer spesialvitenskapenes autonomi: I den klassiske artikkelen "Unity of Science as a Working Hypothesis" (1958) gir Oppenheim og Putnam et klart bilde av den lagdelte verden som jeg har forutsatt i denne oppgaven. Forskjellige spesialvitenskaper har objekter av forskjellig grad av kompleksitet som sitt diskursunivers: Biologer studerer således objekter som er bygget opp av kjemiske objekter og har en kompleksitetsgrad tilstrekkelig for liv. Psykologi er i sin tur studiet av objekter som er bygget opp av biologiske objekter, og som samtidig er tilstrekkelig komplekse til å utvise intensjonalitet.

Anta nå at en vitenskap har nivå L som sitt diskursunivers. Den radikalt positivistiske arbeidshypotesen til Oppenheim og Putnam er at man skal jobbe for redusere de karakteristiske begrepene og lovene på nivå L, til de karakteristiske begrepene og lovene til vitenskapen som har nivået nedenfor (nivå L-1) som sitt diskursunivers. I siste instans vil derfor all spesialvitenskaplig viten prinsipielt kunne utledes fra fysikk. Følgende sitat får frem styrken i denne ambisjonen:

It is not absurd to suppose that psychological laws may eventually be explained in terms of the behavior of individual neurons in the brain; that the behaviour of individual cells – including neurons – may eventually be explained in terms of their biochemical constitution; and that the behavior of individual molecules – including the macromolecules that make up living cells – may eventually be explained in terms of atomic physics. If this is achieved, then psychological laws will have, *in principle*, been reduced to the laws of atomic physics, although it would nevertheless be hopelessly impractical to try to derive the behavior of a single human being directly from his constitution in terms of elementary particles.<sup>123</sup>

Trusselen med dette bildet er nettopp at bl.a. psykologi tilsynelatende kommer innenfor rekkevidde av fysikkens, eller i første omgang nevrofysiologiens, forklaringer.

Vi har sett (1.3) gode pragmatiske grunner for at det ikke kan finnes slike ryddige sammenhenger mellom vitenskapers karakteristiske begreper. Essensielt skyldes dette at

---

<sup>122</sup> For å være litt mer presis så studerer fysikken alle objekter fordi alle objekter har fysiske egenskaper. Det finnes ingen ikke-fysiske *objekter*. Derfor er alle objekter inkludert i fysikkens diskursunivers. Men dette åpner fortsatt for at noen komplekse objekter også kan ha ikke-fysiske *egenskaper*.

<sup>123</sup> Oppenheim & Putnam (1958) s.7.

forskjellige vitenskaper deler opp verden i forskjellige slag, avhengig av bl.a. hvilke erkjennelsesinteresser de svarer til. Vi trenger derfor kanskje ikke ta denne positivistiske drømmen alt for alvorlig. Men selv om de færreste forventer en enhetlig vitenskap *a lá* Oppenheim og Putnam (Putnam gjør det ikke lenger), så er våre metafysiske forpliktelser de samme: Innenfor den mereologiske modellen fikseres makrovitenskaper av mikrovitenskaper. Og den allmenne reduksjonistiske trusselen er – som vi skal se – høyst reell til tross for at positivistiske vitenskapsidealer og reduksjonsmodeller i noen grad har gått av moten. Så selv om den vitenskapelige enhet moderne reduksjonister som Kim er ute etter, er en svakere enhet enn den Oppenheim og Putnam i sin tid søkte, så er psykologiens autonomi på ingen måte garantert.

Jeg har nå motivert trusselen om at den mereologiske modellen av en lagdelt verden truer med å bringe spesialvitenskaper innenfor fysikkens forklaringsmessige rekkevidde, og dermed true disses autonomi. Men som jeg sa innledningsvis, kan vi også finne kimen til en løsning innenfor denne modellen:

For såkalte *emergente egenskaper* er – hvis de finnes – nærmest skreddersydde for å inngå i en slik autonom vitenskap: Grovt og foreløbig kan vi si at de fleste begreper om emergente egenskaper har gått ut på at disse er (i) egenskaper ved komplekse objekter (ii) som er nye i den forstand at de ikke innehaes av objektenes deler, og som (iii) ikke kan forutsies ut fra, eller reduseres til objektets mikrostruktur. Dersom (iii) ikke holder er egenskapen ”resultant.” Som norske språkbrukere bør vi huske at det engelske ordet ”emerge” betyr ”fremtre.” Egenskapene trer først frem på et visst kompleksitetsnivå. Men *at* de fremtrer er et ”brute fact,” det er noe som ikke kan gies en mer fundamental forklaring.<sup>124</sup>

Hvorfor vil det å betrakte psykologiske egenskaper som emergente egenskaper (hvis dette er holdbart) løse motsetningen mellom fiksering og autonomi? Vel, for det første er det at egenskapene er *emergente* egenskaper ved komplekse fysiske/nevrofysiologiske objekter forenelig med at de er *sterkt (mereologisk) superveniente* på mikrostrukturen til disse objektene. (Angående mereologisk superveniens, se (2.4)). Og den ontologiske fikseringen – at de nødvendigvis fremtrer når de subveniente mikrobaserte egenskapen instansieres – innebærer ikke en trussel mot autonomien fordi egenskapene likevel er *nye* i forhold til delegenskapene som er med på å utgjør den mikrobaserte egenskapen. Den epistemologiske fikseringen – at psykologiske fakta er på plass så snart fysiske fakta er på plass – takler

---

<sup>124</sup> Dette premisset kan gies en svakere betydning: Alt vi krever er at instansieringen av den emergente egenskapen i siste instans bunner i at verden *er* slik at egenskapen instansieres. Så hvis noen kan gi en *slags* forklaring som ”rimeliggjør” at egenskapen instansieres, uten å ”nødvendigjøre” instansieringen, så er dette forenelig med emergentisme.



emergentisten ved å si at dette er et "brute fact." At den og den fysiske strukturen instansierer en mental egenskap, er ikke noe som kan gies en ytterligere fysisk forklaring, så det kan eksistere en autonom psykologi om disse egenskapene. Emergentisten kan faktisk også godta bikondisjonaler som tilsier at  $x$  instansierer en mental supervenient egenskap  $M$  hvis og bare hvis den instansierer en av potensielt flere mulige fysiske baseegenskaper  $P_i$ . Noen vil kalle dette en reduksjon, men så lenge det ikke er en *fysisk* historie å fortelle om *hvorfor* bikondisjonalen holder, er det ikke en *forklarende* reduksjon. Det er verdt å merke seg at brolover er (nomologisk) nødvendige, så grunnen til at de ikke er forklarende, er ikke at de kun uttrykker kontingente sammenhenger.

For å generalisere litt, kan vi si at problemet vårt oppsto pga. en konseptuell forskjell eller kløft mellom fysikk og psykologi. Hvis emergentistens strategi holder, så truer ikke denne kløften psykologiens status, til tross for minimal fysikalisme: Det er ikke slik at kløften truer oss til å redusere eller eliminere psykologien. For den konseptuelle kløften svarer til, eller matches av, en ontologisk kløft (de emergente egenskapene er nye egenskaper som ikke innehaes av delene), og av en epistemologiske kløft (*at* de emergente egenskapene instansieres er et "brute fact" ved verden, som ikke kan forklares hverken av fysikk eller av noen annen vitenskap). Emergentisme gir rom for at vesensforskjellige teorier kan beskrive verden riktig.

Det store spørsmålet er selvsagt om det er *berettiget* å legge denne ytterligere strukturen på vårt ontologiske verdensbilde: Vi må ikke bare konstruere et mer presist emergensbegrep som løser konflikten. Dette begrepet må også være konsistent, og det må faktisk finnes emergente egenskaper i verden.

## Del II: Emergentisme

### 5 Emergensbegrepet

#### 5.1 Innledning

Jeg vil her bedrive litt begrepsdesign eller ”conceptual engineering” for å konstruere et emergensbegrep som synes egnet til å løse konflikten som er formulert ovenfor. Underveis vil det også bli klarere hvorfor emergentisme (hvis det er et reelt filosofisk alternativ) løser autonomiproblemet. Emergensbegrepet formuleres innenfor de metafysiske rammene jeg har beskrevet ovenfor: Makroegenskaper er mereologisk superveniente på mikroegenskaper.<sup>125</sup> Termen ”emergens” brukes derfor i forbindelse med egenskapsinstansieringer. Emergente egenskaper er makroegenskaper ved komplekse objekter, og de er derfor mereologisk superveniente på mikrobaserte egenskaper ved objektene.<sup>126</sup> Emergensbegrepet kan generelt sies å ha to komponenter: (i) En *epistemologisk komponent*, som tilsier at den emergente egenskapen i en eller annen forstand er uforutsigbar fra viten om objektets delnivå. (ii) En *ontologisk komponent*, som tilsier at den emergente egenskapen er ”ny” i forhold til objektets delegenskaper: Delene har ikke den emergente egenskapen, og denne egenskapen har kausale krefter som delegenskapene ikke har. Det er derfor en epistemologisk, og en ontologisk kløft mellom mikrovitenskaper og makrovitenskaper som sikrer autonomi. Jeg behandler disse to komponentene suksessivt.

---

<sup>125</sup> Paul Humphreys hevder i ”Emergence, Not Supervenience” at superveniens ikke er egnet til å belyse forholdet mellom det mentale og det fysiske, og at det vi trenger i stedet er emergens. Argumentet synes å bunne i at superveniensrelasjonen også kan gjelde mellom egenskaper på samme nivå, der vi ikke er tilbøyelig til å mene at den superveniente egenskapen er ontologisk sekundær. (Humphreys (1997) s.338-341, spes. s.341.) Men det er ikke klart at denne innvendingen vil berøre mereologisk superveniens. Uansett kan ikke artikkelen sies å gi et argument mot at superveniens faktisk holder. (Det er heller ikke gitt at Humphreys vil si at superveniens ikke holder, selv om han mener at superveniensrelasjonen er ”empty of any scientific content”. (Humphreys (1997) s.237.))

<sup>126</sup> Men lover, hendelser o.a. kan også sies å være emergente, for såvidt som de involverer emergente egenskaper. (Stephan (1992) s.27.) Jeg har karakterisert vitenskaper ut fra de egenskaper de refererer til. Fordi jeg har valgt denne karakteriseringen blir lover etc. *beskrivelsesmessig* sekundære i forhold til egenskaper. Jeg påstår imidlertid ikke noe om hvorvidt det gjelder noe dypere avhengighetsforhold mellom emergente egenskaper og emergente lover, eller omvendt. (Jf. ibid. n.3) Historisk sett har ”emergens” også blitt brukt om evolusjon av stadig mer komplekse objekter. (Se f.eks. Stephan (1992) s.30, som siterer Alexander (1920) s.45.) Jeg behandler ikke dette aspektet ved emergentisme, fordi jeg diskuterer ferdig utviklede objekter og deres egenskaper; ikke objektenes tilblivelse.

## 5.2 Den epistemologiske komponenten

Hvorfor instansierer et komplekst objekt ( $x: [C_1, \dots, C_m; R]$ ) en makroegenskap P? Den epistemologiske komponenten av emergensbegrepet tilsier at dersom P er emergent, så kan det ikke kan gies noe fullstendig svar på dette spørsmålet. La oss generelt si at emergente egenskaper ikke kan forutsies ut fra viten om objektets mikrostruktur. *Resultante* egenskaper *kan* derimot forutsies på denne måten: Et heles temperatur etter en tid kan f.eks. forutsies ut fra egenskaper ved delene (delenes individuelle temperatur og varmekapasitet). Det er klart at denne typen forutsigelse involverer deduksjon gjennom lover (i vårt eksempel termodynamiske lover), og vi kan derfor mer spesifikt snakke om forutsigelse ved *deduksjon*. Det Achim Stephan kaller egenskapslover gir oss et generelt skjema for slik deduksjon, fordi de tilsier at alle objekter med en bestemt mikrostruktur har en bestemt makroegenskap:

For each system  $x$  having the microstructure  $[C_1, \dots, C_n; O]$ :  $x$  instantiates property  $P$  according to a property law  $P_L$ .<sup>127</sup>

(Dersom vi skulle vært helt presise måtte vi også inkludert atomære egenskaper ved delene  $C_1, \dots, C_n$  i  $x$ 's mikrostruktur.) Slike egenskapslover må finnes, fordi superveniensten tilsier at instansieringen av en subvenient egenskap nødvendigvis er tilstrekkelig for instansieringen av den superveniente egenskapen. Det vi må merke oss nå, er at hvis man kjenner et objekts mikrostruktur og den relevante egenskapsloven  $P_L$ , så kan man dedusere at objektet har makroegenskapen  $P$ .

Imidlertid er det klart at  $P$  kan forutsies på to måter her, avhengig av hvordan vi er kommet frem til egenskapsloven  $P_L$ : (i) Hvis  $P_L$  kan deduseres fra vår viten om mikronivået, så er  $P$  ”teoretisk forutsigbar,” (ii) Hvis  $P_L$  bare uttrykker en observert kovariasjon mellom mikrostruktur og makroegenskaper som *ikke* kan deduseres fra vår viten om mikronivået, så er  $P$  derimot ”induktivt forutsigbar.”<sup>128</sup> (At  $P_L$  kun er induktivt *forutsigbar*, innebærer altså – slik jeg bruker begrepet – at instansieringen av  $P$  er ikke-*deduserbar*, i en forstand jeg skal presisere nedenfor. Ideen om ikke-deduserbarhet i emergentisme stammer fra den klassiske emergentisten C.D Broads (1887-1971) emergensbegrep.<sup>129</sup>) En viktig konsekvens av dette er at kun teoretisk forutsigbarhet truer makrovitenskapens autonomi: For dersom  $P$  er teoretisk forutsigbar, så kan vitenskapen om mikronivået gi oss en fullstendig forklaring av hvorfor  $P$  instansieres. Men hvis  $P$  bare er induktivt forutsigbar, så er instansieringen av  $P$  utenfor

---

<sup>127</sup> Stephan (1992) s.33.

<sup>128</sup> Uttrykkene er lånt fra Kim (1999) s.8. Stephan har en alternativ terminologi, der han skiller mellom ante- og post factum forutsigbarhet. (Stephan (1992) s.31)

<sup>129</sup> Broad (1925) s.61.

rekkevidde av mikrovitenskapens forklaringer, og det er derfor rom for en autonom makrovitenskap om P.

Vårt emergensbegrep bør derfor godta induktiv forutsigbarhet, men må utelukke teoretisk forutsigbarhet. På denne måten tillates sammenhenger (supervenienens) mellom makroegenskaper og mikrobaserte egenskaper, men historien om mikronivået kan ikke erstatte historien om makronivået, fordi den mangler forklaringskraft til å forklare de emergente makroegenskapene.

En annen konsekvens av dette er at emergentisten faktisk kan godta bikondisjonaler av typen  $M(x) \leftrightarrow P(x)$ , som forbinder en (f.eks.) psykologisk egenskap med en (f.eks.) fysisk egenskap.<sup>130</sup> (Sterk superveniens og multippel realiserbarhet alene impliserer bare eksistensen av bikondisjonaler som forbinder M med en *disjunksjon* av fysiske egenskaper. (2.3)) Med andre ord kan emergentisten godta muligheten av såkalt nagelreduksjon: Essensen i nagelreduksjon er at en teori reduseres til en annen ved deduksjon. Deduksjonen skjer via såkalte brolover, som postulerer at predikatene i vitenskapen som skal reduseres, har samme ekstensjon som predikatene i vitenskapen man reduserer til. Hvis M er et psykologisk predikat og P er et fysisk predikat, så er altså  $M(x) \leftrightarrow P(x)$  en slik brolov. På grunnlag av slike brolover kan vi skrive en psykologisk lov der M (og en annen mental egenskap) inngår, om til en fysisk lov der P (og en annen fysisk egenskap) inngår. Via brolovene kan så den psykologiske loven utledes fra den fysiske loven.<sup>131</sup> En lignende reduksjonsmodell ligger bak Oppenheims og Putnams arbeidshypotese om en enhetlig vitenskap, som vi så på ovenfor.

Grunnen til at emergentisten kan akseptere nagelreduksjon, er at brolover bare slår fast systematiske sammenhenger mellom egenskapsinstansieringer, de forklarer ikke hvorfor disse sammenhengene finner sted. (Jf. induktiv forutsigbarhet.) Som Kim sier, vil ikke en Nagelsk utledning av psykologi fra fysikk øke vår forståelse av det mentale (MPW 96):

*For it is the explanation of these bridge laws, an explanation of why there are just these mind-body correlations, that is at the heart of the demand for an explanation of mentality. (Ibid.)*<sup>132</sup>

Bare det som kan kalles *forklarende* reduksjoner truer emergentismen.<sup>133</sup>

---

<sup>130</sup> Jeg snakker her om emergentisme generelt, ikke om hva emergentister historisk sett har godtatt.

<sup>131</sup> Se f.eks. Kim (1998b) s.212-13. Denne modellen for reduksjon stammer opprinnelig fra Nagel (1961). (Nagel krevde faktisk ikke at brolovene skulle være bikondisjonaler, og for utledningsformål er selvsagt kondisjonalen  $P(x) \rightarrow M(x)$  tilstrekkelig. (Kim (1998b) s.214) Merk også at egenskapslover bare er kondisjonaler.)

<sup>132</sup> Se også Kim (1999) s.12 og MPW 97 der han i tillegg påpeker at bl.a. enkelte substansdualister også kan godta nagelreduksjon.

<sup>133</sup> Et eksempel på en forklarende reduksjon er reduksjon via identitet: M instansieres samtidig med P fordi  $M=P$ . (En av grunnene til at Kim ønsker identiteter, ikke bare nagelske brolover mellom mentale og fysiske egenskaper er nettopp at identiteter på denne måten er forklarende. (MPW 98.)

Det er dette emergentisten utnytter med sin "brute fact"-strategi. At de emergente egenskapene instansieres når de instansieres er et "brute fact," et grunnleggende trekk ved verden som ikke kan gies en mer fundamental forklaring. Det må bare aksepteres. Det klassiske sitatet i denne sammenhengen er fra emergentisten Samuel Alexander:

The existence of emergent qualities thus described is something to be noted, as some would say, under the compulsion of brute empirical fact, or, as I should prefer to say in less harsh terms, to be accepted with the "natural piety" of the investigator. It admits no explanation.<sup>134</sup>

Jeg har dvellet ved "brute fact"-strategien og den epistemologiske kløften, fordi dette viser at emergentisten kan gi mye til reduksjonisten (nagelreduksjon og induktiv forutsigbarhet) uten å gi slipp på autonomi for spesialvitenskapene.

Men "brute fact"-strategien innebærer *ikke* at fysiske forklaringer aldri er forklaringsmessig relevante for psykologi. (Tvert imot kan nevrofysiologer angivelig si noe om hva som, nevrofysiologisk sett, skiller bevisste fra ubevisste hjerneprosesser.<sup>135</sup> Mer hverdagslig kjenner vi til sammenhengen mellom adrenalinnivå og sinne.) Emergentister som ikke aksepterer dette, ville kjørt seg selv ut på sidelinjen for enhver rimelig debatt om forholdet mellom vitenskapene. Bare det at det gjelder en avhengighetsrelasjon mellom det fysiske og det mentale vil gjøre en slik tese usannsynlig.<sup>136</sup> Det må derfor understrekes at "brute fact" må forstås som "ikke *helt og holdent* deduserbar."<sup>137</sup>

Det må også bemerkes at de emergente egenskapene og egenskapslovene må være ikke-deduserbare i en *absolutt* forstand. Man kan selvsagt formulere et *teorirelativt* emergensbegrep der egenskapene kun er ikke-deduserbare fra én bestemt teori.<sup>138</sup> Dersom denne teorien ikke er ferdig utviklet (og det er de færreste teorier), blir det derfor mulig å hevde at det en gang i en ideell fremtid ikke vil finnes noen emergente egenskaper. Slike teorirelative emergensbegreper kan være nyttige i diskusjon av vitenskapers fremskritt, og det er klart at egenskaper som tidligere ble regnet som emergente, siden har vist seg å være resultante.<sup>139</sup> Imidlertid ville et teorirelativt emergensbegrep utgjøre et uhyre svakt og midlertidig fundament for psykologiens autonomi: Psykologi vil være en egen vitenskap i

<sup>134</sup> Alexander (1920) s.46-47 sitert i f.eks. McLaughlin (1992) s.66.

<sup>135</sup> Hans Flohr (1992) s.224.

<sup>136</sup> Poenget om at avhengighet gjør en slik teori usannsynlig finner vi faktisk allerede hos Broad (1925) s.54. Se også Beckermann (1992) s.100.

<sup>137</sup> Man vil kanskje hevde at spørsmålet om deduserbarhet er et spørsmål om enten-eller. Imidlertid kan vi f.eks. tenke oss at vi en gang i fremtiden kan forstå det mentales superveniensbaser så godt at vi kan gi en teoretisk (ikke induktiv) deduksjon av at *én* mental egenskap instansieres, uten å kunne dedusere *hvilken* egenskap som instansieres.

<sup>138</sup> Stephan ((1992) s.29) tror at Hempel & Oppenheim (1948) var de første til å eksplisitt formulere et teorirelativt emergensbegrep.

<sup>139</sup> Et eksempel er at hvilke kjemiske stoffer som kan binde seg til hverandre en gang ble regnet som emergente fakta. (McLaughlin (1992) s.55). Dette er noe man nå delvis lærer å forutsi teoretisk allerede i ungdomskolen.

dag, men en gang i fremtiden må psykologer kanskje pakke sammen, og overlate det hele til nevrofysiologer, eller kanskje til og med til kjernefysikere.<sup>140</sup> Reduksjonister ville dessuten kunne peke på tidligere vellykkede forklarende reduksjoner av emergente egenskaper. På dette grunnlaget ville de kunne forsvare et reduksjonistisk paradigme for spesialvitenskapelig praksis. Alt i alt er det klart at vi for vårt formål trenger et begrep om absolutt emergens: En måte å oppnå dette på er å inkludere en klausul i definisjonen vår som sier at det ikke eksisterer noen teori  $T$  som egenskapsloven  $P_L$  kan utledes fra.<sup>141</sup> ("Eksisterer" må selvsagt gies en logisk, og ikke spatio-temporal mening.)

Jeg må nå si noe mer presist om deduksjon, og spesielt om hva egenskapslovene skal deduseres *fra*. Hvilken viten om objektets mikrostruktur kan så utgjøre grunnlaget for deduksjonen? For det første er det klart at vi må inkludere noen relasjonelle egenskaper ved delene, i det minste deres romlige organisering. (Beckermann tror at relasjonen mellom delene må beskrives i *utelukkende* romlige termer for ikke å bake inn for mye informasjon i viten om mikronivået.)<sup>142</sup> Dette er klart også fordi vi har sagt at emergente egenskaper er superveniente på en mikrobasert egenskap, og slike egenskaper innebærer også en relasjon mellom delene. ((2.4), MPW 84.) Å inkludere relasjonelle egenskaper er videre nødvendig for i det hele tatt å komme ut av det logiske rom der diskurs om (de isolerte) delene skjer. Dessuten kan den strukturelle organiseringen av delene spille en viktig rolle for helhetens egenskaper: Broad påpeker f.eks. at forskjellige kjemiske forbindelser kan inneholde de samme grunnstoffene i de samme mengdene, så lenge molekylstrukturen er forskjellig.<sup>143</sup> For det andre må vi inkludere viten om delene slik de inngår i andre heler.<sup>144</sup> (Hvilket er rimelig, gitt at det er en del av vitenskapelig praksis å studere *en rekke* objekter eller fenomener og finne sammenhenger mellom dem.) For det tredje er det helt avgjørende at vi har kjenskap til *lover* der delenes egenskaper inngår.<sup>145</sup> Med lover forstår jeg her vitenskapelige verktøy, som postulerer universelle eller statistiske sammenhenger mellom egenskapsinstansieringer. ("E=mc<sup>2</sup>" postulerer således en sammenheng mellom egenskapene masse og energi.) Vi må

---

<sup>140</sup> I praksis er dette antagelig umulig: Mikro og makrovitenskaper er formet av forskjellige pragmatiske hensyn. Det synes rimelig at det psykologien taper i eksakthet og streng lovmessighet i forhold til fysikk får den igjen i det at psykologiske egenskaper er tilgjengelige i praksis, og er relevante for våre (hverdaglige) interesser. (Van Gulick (1993) s.247) Hvis vi mot formodning skulle komme frem til en komplett og forklarende reduksjon av psykologi til fysikk, så ville den reduserte psykologien sannsynligvis vært så ekstremt heterogen, disjunktiv og basert på vanskelig tilgjengelige egenskaper at det ikke ville være hensiktsmessig å erstatte den gamle psykologien. Det jeg diskuterer er ikke autonomi i praksis, men absolutt autonomi.

<sup>141</sup> Slik Stephan (1992) s.39 gjør.

<sup>142</sup> Beckermann (1992) s.103.n.9.

<sup>143</sup> Broad (1925) s.56.

<sup>144</sup> Se Stephan (1992) s. 37, og Broad (1925) s.37.

<sup>145</sup> Beckermann ((1992), s.105) finner dette poenget hos Broad (1925) s.63. Broad tenker her på egenskapslover, men det er klart at poenget holder for lover generelt, f.eks. for lover som brukes til å utlede egenskapslover.

altså ha kjennskap til lover, og ikke bare til egenskaper, dersom vi skal kunne dedusere noe fra egenskapsinstansieringer.

På denne måten kan vi tenke oss at egenskapsloven  $P_L$  – og dermed makroegenskapen  $P$  – kan deduseres fra atomære og relasjonelle egenskaper ved delene, og lover om disse egenskapene ved hjelp av logiske, eller kanskje også matematiske operasjoner.<sup>146</sup>

Imidlertid må vi legge visse føringer på egenskapene og lovene vi deduserer fra. For som Beckermann påpeker må vi utelukke "[...] "ad hoc properties" like the one that components of this kind standing in the relation  $R$  to each other make up a system which has exactly this [emergent] property."<sup>147</sup> Hvis slike egenskaper tillates, kan alle makroegenskaper deduseres, og følgelig blir ingen egenskaper emergente. Han foreslår derfor at bare "basic" eller grunnleggende egenskaper skal tillates, uten å ville presisere hva dette skal bety. En mulig fremgangsmåte ville kanskje være å utelukke egenskaper som ikke er (eller ser ut til ikke å være) teoretisk nyttige i andre sammenhenger. En alternativ tilnærming ville være å ta utgangspunkt i Beckermanns forslag om at relasjonen  $R$  bare skal beskrives i romlige termer. Men vi kan formodentlig ikke si at *alle* relasjonelle ikke-romlige egenskaper er ad hoc. Uansett behøver ikke mangelen på et klart kriterium for grunnleggende egenskaper å utgjøre et problem, for hva som er grunnleggende bør kunne avgjøres ved diskusjon i enkelttilfeller. Jeg forutsetter derfor i det følgende at vi har et slags intuitivt grep om distinksjonen mellom ad hoc og grunnleggende egenskaper.

Beckermann påpeker også at vi må utelukke lignende ad hoc *lover*, som tilsier at alle mikrostrukturer av en bestemt type har den emergente egenskapen, for å unngå et tomt emergensbegrep.<sup>148</sup> Imidlertid har vi sett at egenskapslover har nettopp denne formen, og vi har til og med åpnet for brolover, som jo har en bikondisjonal form. Derfor må vi (som Beckermann) kun utelukke det Broad kaller "unike og ultimate" lover fra deduksjonsgrunnlaget.<sup>149</sup> Broad definerer slike lover negativt ved at de *ikke* oppstår ved å substituere variabler med bestemte verdier i en generell lov. De oppstår heller ikke ved å kombinere to generelle lover. Videre kan lovene bare oppdages ved å studere objekter av en bestemt type, og de lovene gjelder bare for objekter av denne typen.<sup>150</sup>

---

<sup>146</sup> Det er selvsagt mulig at det som utledes ikke er *helt* isomorft med teorien som skal reduseres, men innebærer en form for korreks til den. (Se f.eks. Churchland & Churchland (1990) s.70-71.)

<sup>147</sup> Beckermann (1992) s.104.

<sup>148</sup> Ibid. s.105.

<sup>149</sup> Ibid. s.106.

<sup>150</sup> Broad (1925) s.65. Broad snakker om [bro]lover være emergente. (Se f.eks. Broad (1925) s.72.) Men som nevnt kan vi oversette dette til at makroegenskapen som er involvert i loven er emergent. Broad's begrep om "ultimate characteristics" (ibid. s.78) svarer forøvrig til "emergente egenskaper."

Broads eksempel er at det (ifølge ham selv, vel å merke) ikke finnes noen generelle lover som gir sammenhenger mellom egenskaper klor og sølv har alene, eller i andre forbindelser, og egenskaper som sølvklorid har. Sølvklorids egenskaper finner vi bare ved spesifikke studier av substansen sølvklorid.<sup>151</sup> Sølvklorideksempelen gir oss altså en del hele sammenheng som er ”brute,” eller i Broads terminologi *ultimat*. Derfor kan den også bare avdekkes induktivt. Men et nytt element er kommet til, nemlig *unikhet*. I dette ligger at egenskapslover for emergente egenskaper er *partikulære*, og ikke *generelle* lover. De gir derfor ikke generell viten. Vår viten om sammenhengen mellom emergente egenskaper og deres superveniensbaser er derfor alltid en viten om *enkelttilfeller*: Sammenhengen mellom deegenskaper og heleits egenskaper i sølvklorideksempelen er for det første en viten som bare gjelder sølvklorid, og for det andre en viten som bare kan oppdages ved å studere substansen sølvklorid.

Dermed sikrer emergentisme autonomi for makrovitenskapene av *to* grunner: (i) Fordi egenskapslovene er *ultimate* (dvs. ”brute facts”), men også (ii) fordi de er *unike*, dvs. de forteller oss bare om partikulære sammenhenger mellom mikrobaserte egenskaper og emergente makroegenskaper. Det er selvsagt logisk mulig at et lite antall ryddige, men ultimate lover forbinder mentale og fysiske egenskaper. Men hvis situasjonen er som Broad skisserer, så finnes det tvert imot et *stort* antall lover, som ikke bare er ultimate, men også unike. Dette ville utvide den epistemologiske kløften mellom psykologi og fysikk, og dermed ytterligere sikre psykologiens autonomi. Dette er som vi skal se et mulig utgangspunkt for en kritikk av Kim. (6.4)

### 5.3 Den ontologiske komponenten

Emergente egenskaper skiller seg ontologisk fra mikroegenskapene de er superveniente på, ved at de utgjør et *eget* område eller nivå av virkeligheten som en autonom makrovitenskap kan studere. Men dette nivået er selvsagt avhengig av rent fysiske nivåer for sin eksistens, slik minimal fysikalisme forutsetter. Dessuten har objektene på dette nivået også fysiske egenskaper og kan følgelig også studeres *qua* fysiske objekter. Det er de emergente *egenskapene* ved disse objektene, eller vitenskapens *tema* om man vil, som utgjør det autonome området.

---

<sup>151</sup> Broad (1925) s.64. Se også Beckermann (1992) s.105.



Jeg formulerer dette som at enhver emergent egenskap  $P(x)$  er *ny* i forhold til egenskaper ved mikrostrukturen til  $x$ :  $[C_1, \dots, C_m; R]$ .<sup>152</sup> For det første er  $P$  *ny* i en nokså triviell forstand: Siden  $P$  er en makroegenskap er det ingen av delene  $C_1, \dots, C_m$  som har  $P$ . Men dette er tilfellet også for *resultante* makroegenskaper: En bestemt ansamling vannmolekyler har f.eks. ved en bestemt temperatur og et bestemt trykk et volum på 1 liter, og dette er en egenskap ingen av de individuelle vannmolekylene i ansamlingen har. Men molekylene fyller også et (riktignok mindre) volum, så egenskapen å ha et volum på 1 liter er ikke forskjellig fra delegenskapene. Tvert imot er den helt og holden redusibel og resultant. Den romlige relasjonen molekylene er organisert i, er heller ikke ”ny” i noen sterk forstand.

Men for det andre er vår emergente  $P(x)$  *ny* forsåvidt som den er en egenskap av en annen *art* enn mikroegenskapene. Den oppfører seg på en annen måte, og beskrivelser av den forutsetter andre prinsipper. Dette harmonerer med at psykologi og fysikk er vesensforskjellige begrepslige praksiser.

Mer konkret har jeg sagt at for at psykologien skal ha realitet, så må dens begreper referere til egenskaper med kausale krefter (Prinsippet om det mentales realitet, (3.3)): I tråd med dette koker det at emergente egenskaper er *nye*, ned til at de har kausale krefter som delegenskapene ikke har.<sup>153</sup> Med ”nye” må vi her som vi har sett forstå at de adlyder andre prinsipper, f.eks. psykologiske prinsipper. Novelitet er nært knyttet til kausal realitet. At realitet forutsetter kausal potens, er et velkjent poeng blant emergentister. Dette kommer særlig godt frem av følgende berømte sitat om epifenomenalisme fra Alexander. Kim synes selvsagt at dette er ”a marvellous paragraph”:

The doctrine [...] supposes something to exist in nature which has nothing to do, no purpose to serve, a species of *noblesse* which depends on the work of its inferiors, but is kept for show and might as well, and undoubtedly would in time be abolished.<sup>154</sup>

For å oppsummere så har emergentister følgende forpliktelser: (i) Visse komplekse objekter har kausale krefter som objektenes deler ikke har. (ii) Disse kreftene adlyder autonome prinsipper. Jeg tror (i) og (ii) kan fanges opp i det McLaughlin kaller *konfigurasjonelle krefter*. Disse er fundamentale (og følgelig irreducibile) krefter som bare kan utøves av bestemte konfigurasjoner av bestemte partikler, og ikke av par av

<sup>152</sup> Man kunne selvsagt også si at den *epistemologiske* komponenten handler om novelitet: Emergente egenskapene er nye fordi de ikke kan reduseres forklarende til baseegenskapene. Men denne noveliteten er *grunnet* i den ontologiske noveliteten: Irreducibiliteten skyldes at verden er innredet på en bestemt måte. Termen ”novelty” brukes forøvrig ofte i emergensdiskusjoner, f.eks. i Stephan (1992) s.30-32.

<sup>153</sup> Kim vektlegger at emergentister har denne forpliktelsen, og hans kritikk går ut på å vise at de ikke kan innfri den. (Kim (1999) s.25.)

<sup>154</sup> Alexander (1920) s.8, Kim (1993d) s.202.

elementærpartikler.<sup>155</sup> Grovt sett er konfigurasjonelle krefter altså krefter ved spesielle typer komplekse objekter. Ut fra dette kan vi si at f.eks. strukturer som utviser mentalitet har konfigurasjonelle krefter, som ikke-mentale strukturer mangler. Til tross for at McLaughlin mener at det empirisk sett har vist seg at det ikke finnes noen konfigurasjonelle krefter,<sup>156</sup> så argumenterer han for at de ikke er inkonsistente med fysikkens verdensbilde.<sup>157</sup> Essensen i argumentet synes å ligge i at lover som omhandler fysiske egenskaper er generelle, og at så lenge lovene som omhandler spesielle konfigurasjoner ikke er i konflikt med de generelle lovene, så har vi ikke noen kontradiksjon.

For å ta et av hans eksempler, så er det en *generell* lov at summen av alle kreftene som virker på et legeme er lik produktet av legemets masse og akselerasjon. ( $\Sigma F=ma$ .) Men så lenge dette generelle kravet til summen av kreftene oppfylles, kan noen av kreftene meget vel være konfigurasjonelle.<sup>158</sup> Dette innebærer altså at noen av kreftene som virker på legemet kan være regulert av lover eller prinsipper som ikke gjelder for alle legemer.

En annen måte å si dette på vil kanskje være å knytte McLaughlins poeng til det at fysikken er en generell vitenskap: Fysikken er generell, for såvidt som alle objekter har fysiske egenskaper. Følgelig er fysikk studiet av objekter av alle mulige slag. Fysikk er også generell i den forstand at den formulerer lover som gir generelle sammenhenger mellom objekters kausale krefter som må gjelde for alle objekter. Men at fysikk er generell i den sistnevnte forstanden, betyr at den abstraherer fra hvilke objekter som utøver kreftene, siden alle objekter som overhodet utøver krefter må tilfredsstille fysiske lover. *Derfor* kan enkelte av kreftene som virker være konfigurasjonelle, dvs. bare utøves av bestemte typer objekter.

Dette er uhyre viktig for min problemstilling: Motsetningen mellom fysikk og psykologi ligger i at de har forskjellige prinsipper for forklaring, den er ikke en *logisk* motsetning. I den grad psykologi står i en kontradiksjonsrelasjon til fysikk, så må vi revurdere psykologiens sannhetsverdi.<sup>159</sup> Vi forventer at mentale egenskaper kan forandre verdens gang, men vi ønsker ikke en psykologi som tilsier at de (f.eks.) kan oppheve tyngdekraften!

---

<sup>155</sup> McLaughlin (1992) s.52

<sup>156</sup> Ibid. s.54-55, 75 og 90-91.

<sup>157</sup> Se ibid. s.74 for denne konklusjonen.

<sup>158</sup> Ibid. s.70-71

<sup>159</sup> Eller *fysikkens* sannhetsverdi. Dette er en logisk mulighet, på samme måte som man konsistent kan forkaste substansdualisme til fordel for en monisme der substansen er en sjelelig substans (*res cogitans*) og ikke en materiell substans (*res extensa*).

## 5.4 "Downward causation" er fundamental

Hva slags kausale relasjoner er emergentisten dermed forpliktet til? Innenfor ontologien av en verden som er delt inn i nivåer etter den mereologiske relasjonen, er det klart at en egenskap E kan forårsake en annen egenskap F på tre måter, avhengig av hvilket nivå F befinner seg på: Dersom F er på samme nivå som E, har vi det Kim kaller "*same-level causation*", dersom F er på et lavere nivå, har vi "*downward causation*" og hvis F er på et høyere nivå, har vi "*upward causation*".<sup>160</sup> (Jeg er beklageligvis ikke i stand til å finne en god norsk oversettelse av "*downward causation*," og bruker heretter ofte forkortelsen DC i stedet.)<sup>161</sup> Kim argumenterer flere steder for at DC er fundamental, fordi DC *forutsettes* av de to andre tilfellene av "*causation*."

Kims argument kan oppsummeres slik: Anta at vi har et tilfelle av "*same-level causation*," dvs. anta at en mental egenskap M (eller mer generelt en egenskap M) forårsaker M\*, som er på samme nivå som M. (MPW 41.) Et eksempel på dette vil formodentlig være at en tanke forårsaker en annen. Minimal fysikalisme tilsier at M\* har en fysisk superveniensbase P\*. Kims poeng er at vi nå har to *konkurrerende* forklaringer av M\*'s instansiering: For vi kan forklare M\* enten med henvisning til dens superveniensbase P\*, eller med henvisning til dens antatte årsak M. Hva mer er: Forklaringen som viser til P\* nyter en spesiell status: For det første er P\* *tilstrekkelig* for M\* uavhengig av om M instansieres eller ikke. For det andre er P\* (eller en annen superveniensbase) *nødvendig* for M\*. (Alt dette følger fra egenskaper ved superveniensrelasjonen.) Kim konkluderer med at hvis M skal forårsake M\*, så må den gjøre det *gjennom å forårsake P\**, dvs. gjennom DC. (MPW 42.) "*Same-level causation*" forutsetter altså DC.<sup>162</sup>

Konklusjonen kan tilsynelatende unngås ved å benekte at forklaringene er konkurrerende (dvs. benekte at det som Kim sier er "a real tension" mellom dem (MPW 42)), og insistere på at vi rett og slett har *to* forklaringer av M\*. Hansen foreslår således at noen kanskje kan benekte at det er en spenning mellom forklaringene, ved å si at de besvarer forskjellige spørsmål: Forklaringen ved M er en kausalforklaring, mens forklaringen ved P\* er en del av "a *constitutive story*."<sup>163</sup> Hansens (aristoteliske?) forslag synes for meg svært plausibelt, og det harmonerer med at våre forklaringer av ett og samme eksplanandum ofte

---

<sup>160</sup> Kim (1999) s.22

<sup>161</sup> Charbel Ninō El-Hani & Claus Emmeche benytter faktisk samme forkortelse. (El-Hani & Emmeche (2000) s.234.)

<sup>162</sup> Argumentet og andre formuleringer av prinsippet finnes også i f.eks. Kim (1998b) s.230-231 eller (1999) s. 23-24. I sistnevnte finnes et tilsvarende argument for "*upward causation*."

<sup>163</sup> Hansen (2000) s.465-466.

svarer til forskjellige pragmatiske hensyn: Psykologer og nevrologer kan således tenkes å studere samme mentale tilstand, men i henholdsvis kausale og konstitutive eksplanatoriske kontekster. De vil forklare tilstanden ut fra forskjellige pragmatiske hensyn.

Kim vil likevel ikke kjenne seg berørt av denne kritikken: For å generere spenningen mellom forklaringene forutsetter han nemlig et interessant prinsipp, som han gjør helt eksplisitt et annet sted:

The general principle of explanatory exclusion states that two or more complete and independent explanations of the same event or phenomenon cannot coexist.<sup>164</sup>

Prinsippet om eksplanatorisk eksklusjon utelukker altså at det kan eksistere to komplette og uavhengige forklaringer av  $M^*$  *selv om* de to forklaringene er av forskjellig natur. Pga. prinsippet må Kim spørre seg om hvilken relasjon det er mellom forklaringene ved  $M$  og  $P^*$ : For en konsekvens av prinsippet må være at når vi har to forklaringer av samme eksplanandum som pretenderer å være komplette og uavhengige, så har vi kun følgende muligheter: (i) Vise at forklaringene likevel *er* avhengige, (ii) vise at de *ikke* er komplette, eller (iii) *eliminere* en (eller begge) av dem. I denne sammenhengen velger Kim altså (i):  $M$  og  $P^*$  er *ikke* uavhengige forklaringer av  $M^*$ , fordi  $M$  forårsaker  $M^*$  gjennom å forårsake  $P$ .

Bør vi akseptere Kims konklusjon? Gitt mine sympatier for Hansens forslag er jeg ambivalent til prinsippet Kim bruker for å få konklusjonen. Kim antyder at det å benekte prinsippet innebærer en slags instrumentalisme om forklaring.<sup>165</sup> Det er vanskelig å se at dette er riktig for prinsippet i den *generelle* formuleringen vi har sett ovenfor: Forskjellige typer forklaringer – spesielt kausale og konstitutive – kan plukke ut forskjellige trekk ved verden, og således være forenelige med realisme om forklaring. Hvis man derimot snevrer inn prinsippet til å gjelde kun for kausalforklaringer, så blir situasjonen annerledes: For i forbindelse med kausalforklaring synes realisme om forklaring å kreve at det finnes høyst en komplett forklaring. I dette spesialtilfellet vil jeg derfor akseptere prinsippet: Her svarer forklaringene til samme type spørsmål, og Hansens innvending gjelder ikke.<sup>166</sup> (Dette blir viktig for Kims argumentasjon senere.) Kim har forøvrig også et mer pragmatisk argument for prinsippet, nemlig at formålet med forklaringer er å *forenkle*, og at flere forklaringer av

---

<sup>164</sup> Kim (1989a) s.250. Kim refererer riktignok ikke helt eksplisitt til prinsippet i MPW. Imidlertid synes argumentasjonsstrukturen hans i MPW å bekrefte at han forutsetter prinsippet. At prinsippet spiller en avgjørende rolle for Kim vil bli begrunnet nærmere i diskusjonen av "superveniensargumentet" (7.3), der Kim argumenterer for at fysiske årsaker ekskluderer muligheten av irreducible mentale årsaker. (MPW 43-45.)

<sup>165</sup> Kim (1989a) s.250.

<sup>166</sup> Hans innvending er da heller ikke rettet mot dette spesialtilfellet. Kim har forøvrig ikke med prinsippet utelukket at enkelttilfeller av kausal overbestemmelse kan forekomme: For han antyder nemlig at hvis en hendelse har to tilstrekkelige årsaker, så må en komplett forklaring vise til begge årsakene. (Kim (1989a). s.252)

samme eksplanandum ikke forenkler.<sup>167</sup> Dette argumentet kan kanskje ha en viss apell i forbindelse med *kausal*forklaringer. Imidlertid kan vel heller ikke dette holde for forklaringer generelt.

Når dette er sagt, tror jeg at man også uavhengig av prinsippet må innrømme at det er en *slags* spenning mellom forklaringene: For minimal fysikalisme forplikter oss som vi har sett til at P\* – også *uten* M – er tilstrekkelig og nødvendig for M\*. Så P\* (eller en annen fysisk egenskap) truer faktisk med å gjøre M overflødig. Å si at M forårsaker M\* gjennom å forårsake P, kan kanskje være den eneste måten å gi M en eksplanatorisk relevant rolle på. Poenget her er altså ikke at P\* gir en annen type forklaring, men at den gir en *tilstrekkelig* forklaring.

Kim nevner også at man kan akseptere hans konklusjon også uten å se noen "tension" mellom forklaringene, dersom man aksepterer følgende prinsipp om DC: "*To cause a supervenient property to be instantiated, you must cause its base property (or one of its base properties) to be instantiated.*" (MPW 42.) Alt i alt synes Kims konklusjon rimelig: DC er fundamental. Men også uavhengig av om dette argumentet er holdbart, må det innrømmes at DC er viktig: Store deler av vårt mentale liv har kroppslige bevegelser som sitt mål.

Fra Kims argument følger altså tilsynelatende at det ikke er nok å si at enhver forandring på makronivå vil *implisere* en forandring på mikronivå, kanskje som et biprodukt. Man må også hevde at: Enhver forandring på makronivå må skje *via* en forandring på mikronivå. Dette betyr at emergentisten er forpliktet til DC, hvis emergente egenskaper skal være nye og reelle. *Emergente egenskaper må være årsaker som virker nedad fra emergent makroegenskap til mikrobasert egenskap, fra supervenient (betinget) til subvenient (betingende) nivå.*

---

<sup>167</sup> Kim (1989a) s.254. Mer detaljert argumentasjon finnes også i denne artikkelen.

## 5.5 Formulering av emergensbegrepet

Det som er sagt om emergensbegrepets epistemologiske og ontologiske komponent, kan oppsummeres i følgende definisjon:

La  $x: [C_1, \dots, C_m; R]$  være et komplekst objekt slik at  $P_1(C_1), \dots, P_m(C_m)$ , og la  $P(x)$  være en makroegenskap.

Da er  $P$  *emergent* hvis og bare hvis:

- (i) Det finnes en egenskapslov  $P_L$  som sier at alle objekter med mikrostruktur som  $x$  instansierer  $P$  med nomologisk nødvendighet.
- (ii) Det ikke finnes noen teori  $T$  som  $P_L$  (og dermed  $P(x)$ ) kan deduseres fra, der  $T$  omfatter viten om  $C_1, \dots, C_m$  (a) slik de inngår i  $x$  og i andre objekter, (b) deres *grunnleggende* relasjonelle og ikke-relasjonelle egenskaper og (c) lover som som *ikke er unike og ultimate*.
- (iii)  $P$  kan utøve DC, dvs.  $P(x)$  kan være årsak til egenskapsinstansieringer på mikronivået som  $C_1, \dots, C_m$  befinner seg på.

Hvis (ii) ikke holder, er  $P$  *resultant*. Hvis noen psykologiske egenskaper tilfredsstiller denne definisjonen, så er problemet om psykologiens status løst.

## Del III: Kims kritikk av emergentisme

### 6 Utgangspunktet for et forsvar mot Kims kritikk av emergentisme

#### 6.1 Innledning

Jeg vil i dette kapittelet kort ta opp hva slags problemer emergentisten står ovenfor, og spesielt diskutere Kims kritikk at emergensbegrepets epistemologiske komponent.

#### 6.2 Innledende om problemer for emergentisme

Før jeg kommer inn på Kims kritikk av emergentisme, må jeg kort si noe om to innledende problemer eller innvendinger mot emergentisme: Én trussel mot emergentisme er det vi kan kalle *det empiriske problemet*: McLaughlin hevder som vi har sett (5.3) at selv om emergentisme er en konsistent doktrine, så har det empirisk sett vist seg at det ikke finnes emergente egenskaper med kausale krefter. Dette begrunner han med at kvantemekanikk og annen vitenskapelig utvikling har gitt ”compelling evidence” for at det ikke finnes konfigurasjonelle (f.eks.) kjemiske krefter.<sup>168</sup> Kim legger formodentlig noe av det samme i sitt eksempel om at vanns gjennomskinnelighet en gang ble regnet for en emergent egenskap, men siden har vist seg å være en resultant egenskap.<sup>169</sup>

Til dette problemet vil jeg bare bemerke tre ting: (1) For å dokumentere fraværet av emergente egenskaper er det ikke tilstrekkelig å vise til reduksjoner mellom vitenskaper som kjemi og fysikk. Man vil *forvente* at fysikk kan forklare mange kjemiske egenskaper, nettopp fordi disse vitenskapene er så like med hensyn til pragmatiske interesser og prinsipper for begrepsanvendelsen. Den konseptuelle kløften som skiller fysikk fra kjemi, er mye mindre enn kløften som skiller nevrobiologi og intensjonal psykologi.

(2) Nevrofysiologen og nobelprisvinneren Roger Sperry (1913-1994) hevder at flere vitenskapsmenn faktisk arbeider ut fra emergentistiske forutsetninger. Han mener at det har foregått et paradigmeskifte fra mikroreduksjonisme til en form for emergentisme der makroegenskaper utøver genuin kausal påvirkning på mikronivået.<sup>170</sup> Jeg kommer tilbake til Sperry og dette påståtte paradigmeskiftet mot slutten av oppgaven. (8.3) Hvis Sperry har rett,

---

<sup>168</sup> McLaughlin (1992) s.89.

<sup>169</sup> Kim (1998b) s.227.

<sup>170</sup> Se spesielt Sperry (1991).

kan McLaughlins innvending bunne i en misforståelse av hvilke forutsetninger vitenskapsmenn arbeider ut fra.

(3) Faktisk er det plausibelt at det finnes i det minste noen emergente egenskaper: For det ser ut til å herske relativt stor enighet om at det er et "explanatory gap" mellom nevrologisk viten og viten om såkalte *qualia* eller subjektive, rå følelser.<sup>171</sup> Man kan ikke utlede en følelses kvalitative natur fra viten om hva som skjer i nervesystemet når følelsen finner sted. *Qualia* synes altså å være gode eksempler på emergente egenskaper. Det er derfor ikke overraskende at Broad regnet loven som forbinder kjemisk viten om ammoniakk med lukten av ammoniakk for emergent.<sup>172</sup> Stephan sammenligner faktisk Broads argument her med Frank Jacksons "knowledge argument".<sup>173</sup> Likeledes snakker Paul M. Churchland om "emergent qualia" i forbindelse med Jackson.<sup>174</sup> Men for at problemet i denne oppgaven skal ha en emergentistisk løsning, må selvsagt *intensjonale*, og ikke bare fenomenale egenskaper, være emergente. Jeg nevner *qualia* hovedsaklig for å få frem at det ikke er så totalt urimelig at det skulle finnes emergente egenskaper.

Et annet mulig problem er at emergentisme, med sin "brute fact"-strategi rett og slett er uforenelig med den forutsatte realismen om forklaring. Denne innvendingen koker ned til at det å insistere på at verdens innredning ikke kan gies noen ytterligere forklaring, ikke kan kalles en tilbakeføring til trekk ved verden. Men dermed er den også en misforståelse: Realisme om forklaring forutsetter ikke at forklaringer alltid må grunnes i andre forklaringer i det uendelige. Før eller siden bunner de i at verden *er* slik. Noen vil hevde at vi kun må forvente et lite antall enkle "brute facts" i verden. Dersom instansieringen av mentale egenskaper er "brute facts," så har vi derimot et stort antall uforklarlige fakta i vår ontologi.<sup>175</sup> Men denne forventningen synes å forutsette det fysikalistiske idealet om en sparsom og enkel ontologi. En fruktbar beskrivelse av antireduksjonisme er derimot som teoretisk pluralisme. Verden er rett og slett beskrivbar av en rekke teorier. Gitt *denne* forutsetningen vil det være rimelig å forvente at forsøk på å bringe psykologiske og fysiske egenskaper i ryddig korrespondanse med hverandre bare resulterer i en rekke "brute facts." Diskusjonen om hvor mange "brute facts" som finnes, og hvor i verden man finner dem, synes å være mer preget av intuisjoner og estetiske fordommer, enn av reelle argumenter.

---

<sup>171</sup> Enigheten er likevel ikke absolutt, se f.eks. Churchland (1985) s.25-28, (1995) s.205.

<sup>172</sup> Broad (1925) s.71-72.

<sup>173</sup> Stephan (1992) s.41.n.17. "The knowledge argument:" Se f.eks. Jackson (1986) s.291.

<sup>174</sup> Churchland (1985) s.25.

<sup>175</sup> Kim nevner et lignende argument i forbinelse med en diskusjon av type-type identitetsteorien. (Kim (1998b) s.54)



Når det er sagt, må det innrømmes at antagelsen om at det kun er et lite antall "brute facts" til en viss grad kan rettferdiggjøres metodologisk: Det å søke etter et lite antall generelle lover har f.eks. vist seg å være svært fruktbar i naturvitenskapen. Det å anerkjenne noe som et "brute fact" setter derimot en stopper for videre undersøkelse. Likevel: Hvis noe motstår gjentatte og iherdige forsøk på forklaring, så må man ikke la estetiske preferanser alene hindre en i å erklære det for et "brute fact."

### 6.3 Generelt om Kims kritikk av emergentisme

Jeg vil bruke emergensbegrepet jeg nettopp har formulert til å forsvare emergentisme mot Kims kritikk. Det må bemerkes at ikke all denne kritikken er eksplisitt rettet mot emergentisme. Imidlertid kan Kims reduktive posisjon sies å *utgjøre* en kritikk mot det jeg har kalt emergensbegrepets epistemologiske og ontologiske komponent. Epistemologisk hevder han at funksjonalistiske lokale reduksjoner kan besvare "the explanatory question," dvs. forklare hvorfor visse objekter har mentale egenskaper, og dermed lukke den epistemologiske kløften. Ontologisk argumenterer han mot muligheten av DC fra ikke-reduisible egenskaper.

Nedenfor (6.4) vil jeg se på Kims lokale reduksjonisme. Han mener, som vi skal se, at dersom mentale egenskaper kan funksjonaliseres, så kan de og deres oppførsel *forklares* ut fra fysisk teori. Men gitt multipl realisérbarhet kan de bare reduseres lokalt, dvs. relativt til bestemte strukturer. Jeg vil argumentere for at Kim ikke kan besvare "the explanatory question" hvis den fysiske teorien hans blir så lokal at den inneholder unike lover i Broads forstand. Problemet er at for vanlig funksjonalisme, er det liten grunn til å tro at den lokale teorien må inneholde slike lover, og Kims reduktive prosjekt lykkes i dette tilfellet. Så det vil vise seg at emergentisme må være *en sterk form for egenskapsdualisme* for å unngå å dele skjebne med funksjonalisme. Denne egenskapsdualismen innebærer dessuten at det at de emergente egenskapene har egne kausale krefter, og evnen til DC blir en essensiell forutsetning for emergentisten: Emergentisten har altså et svar på kritikken av den epistemologiske komponenten, bare hvis han har et svar på kritikken av den ontologiske komponenten.

Når det gjelder den ontologiske komponenten, så har Kim utrettelig angrepet DC gjennom en rekke artikler. Disse angrepene kan sies å utgjøre en hovedpilar i hans reduktive prosjekt: Essensielt ønsker han å vise at antireduksjonister er forpliktet til DC, og at DC er umulig, slik at antireduksjonisme blir en håpløs posisjon. Hans hovedargumentasjon mot DC

har vært eksklusjonsbasert: Han vil vise at et objekts fysiske egenskaper gjør alt det kausale arbeidet, og at dets mentale egenskaper dermed utelukkes (ekskluderes) som årsaker. Dette argumentet finner sitt mest modne uttrykk i MPW,<sup>176</sup> og jeg vil diskutere det i (7.2-7.6) Men nylig<sup>177</sup> har han også kommet med et argument mot DC som bygger på den asymmetriske avhengighetsrelasjonen som minimal fysikalisme forutsetter: Dette argumentet kan sies å skulle vise at DC så å si ”slår bena under seg selv,” ved å forsøke å endre sin egen betingelse. Jeg vil diskutere dette argumentet i (7.7). Men først vil jeg altså diskutere Kims kritikk av emergensbegrepets epistemologiske komponent.

#### 6.4 Kims kritikk av den epistemologiske komponenten

Mye av kraften bak multipel realiserbarhetsargumenter for antireduksjonisme ligger i at det ikke finnes ryddige sammenhenger mellom spesialvitenskapenes slag (”kinds”) og fysikkens slag.<sup>178</sup> Det synes derfor å være en sammenheng mellom disse argumentene og Broads begrep om en unik lov, dvs. en lokal og partikulær sammenheng mellom mikrostruktur og makroegenskap. Jeg tror faktisk at begrepet om unike (og ultimate) lover gjør mye av den samme argumentative jobben i Broads posisjon, som multipel realiserbarhetsargumenter gjør i nyere antireduktive posisjoner.

Samsvaret mellom emergentisme og slike argumenter er relevant for en kritikk av Kims posisjon, fordi den reduksjonsmodellen Kim nå går inn for er en variant av det han kaller ”species-restricted, or ”local,” reductions” (MPW 94.) Et eksempel vil være at smerte (som kan realiseres av forskjellige tilstander i forskjellige arter), reduseres til den passende tilstanden i hver enkelt art eller system. (MPW 93.) Kim oppsummerer sin tretrinns reduksjonsmodell karest i en ny artikkel: (1) Definér den superveniente egenskapen som skal reduseres ut fra den kausale rolle den spiller i forhold til andre egenskaper (dvs. at egenskapen funksjonaliseres). (2) Finn baseegenskapen som realiserer den kausale rollen *i det gitte systemet*. (3) Finn en teori på det subveniente nivået som forklarer *hvorfor* denne egenskapen tilfredsstiller den kausale rollen.<sup>179</sup> Han mener videre at egenskapen som skal reduseres, kan identifiseres med den realiserende egenskapen, *relativt* til den organismen/systemet det er tale om.

---

<sup>176</sup> Men se også f.eks. Kim (1989b).

<sup>177</sup> Kim (1999) og (2000).

<sup>178</sup> Se f.eks. Fodor (1974) s.144-145 for denne konklusjonen.

<sup>179</sup> Se Kim (1999) s.10-11. Her snakker han også mer generelt om å finne alle realisatorene, men siden han utelukker at en disjunksjon av egenskaper er en egenskap (MPW 106-109), kan han ikke sette likhetstegn mellom egenskapen og disjunksjonen av realisatorer. Kim beskriver også reduksjonsmodellen sin i MPW 24-25 & 98-99.

Funksjonaliseringen i trinn (1) innebærer at den funksjonaliserte egenskapen blir en såkalt annenordens egenskap, som Kim definerer slik:

*F is a second-order property over set **B** of base (or first-order) properties iff F is the property of having some property P in **B** such that D(P), where D specifies a condition on members of **B**. (MPW 20.)*

For funksjonelle egenskaper er selvsagt D et krav om at P skal spille en bestemt kausal rolle. Dette er avgjørende for Kims begrunnelse i MPW for å identifisere den funksjonelle egenskapen med dens realiserende egenskap. Begrunnelsen kan kanskje sies å være todelt: (i) Fra måten annenordens egenskaper er definert på, følger at de er identiske med egenskapene egenskapene som realiserer dem. For la oss si at F er egenskapen å ha baseegenskapen Q i et gitt system. Da er  $F=Q$ . For som Kim sier: "But in general the property of having property Q = property Q." (MPW 99.) Den dypere grunnen til denne identiteten er at annenordens egenskaper formuleres ved logiske operasjoner på egenskaper i basemengden. Og man kan ikke generere nye uavhengige egenskaper ved slike logiske operasjoner. (Dersom det gir mening å snakke om konjunktive egenskaper, så har vi et mulig unntak.) Siden funksjonelle egenskaper oppstår ved slike logiske operasjoner, så er de strengt tatt bare konseptuelle "egenskapsdesignatorer." De er begreper, som brukes til å si at et objekt har en baseegenskap som fyller en bestemt kausal rolle. De er ikke egenskaper i fullblods forstand. (MPW 103-104.)

(ii) Mer metafysisk bruker Kim sitt "causal inheritance principle" for å begrunne identifiseringen:

If a second-order property *F* is realized on a given occasion by a first-order property *H* (that is, if *F* is instantiated on a given occasion in virtue of the fact that one of its realizers, *H*, is instantiated on that occasion), then the causal powers of this particular instance of *F* are identical with (or are a subset of) the causal powers of *H* (or of this instance of *H*). (MPW 54.)

Funksjonelle egenskaper (*qua* annenordens egenskaper) har altså ingen kausale krefter som skiller dem fra realisatorene, og Kim slutter fra denne mangelen på kausal forskjell til at de er identiske med den baseegenskapen som realiserer dem i den relevante strukturen (MPW 110.)

Dersom vi godtar at logiske operasjoner på egenskaper ikke skaper nye egenskaper, og at forskjellige egenskaper må ha forskjellige kausale krefter, så ser det ut til at konklusjonen følger: Funksjonalisme (formulert ut fra annenordens egenskaper) impliserer lokal reduksjonisme. Imidlertid kan emergentistiske ideer kanskje brukes til en kritikk av Kims reduksjonisme her.

Kim forutsetter nemlig at han har en subvenient teori som forklarer *hvorfor* den funksjonelle egenskapen realiseres av den fysiske baseegenskapen. Men gitt at denne teorien

er lokal, er det mulig at den kan avvises som ikke-forklarende fordi den inneholder en unik lov. I så fall er Kims reduksjonisme, til tross for at den gir lokale identiteter, ikke *forklarende*, fordi dens forklaring av de lokale identitetene ikke er *allmenn* i sitt vesen.<sup>180</sup> For hans historie om ”hvorfors” vil inneholde lover som bare gjelder for dette systemet, og som bare kan oppdages ved studier av dette systemet. Kort sagt; *hvis* det finnes unike lover i teorien, så har ikke Kim besvart ”the explanatory question.”

Kim har ikke én fysisk historie å fortelle om psykologi: Psykologi – en relativt enhetlig vitenskap – splittes tvert imot opp i mange lokale, partikulære og *heterogene* fysiske historier. Denne trusselen mot lokale reduksjoners forklaringskraft er også større enn Kims beskrivelse kan gi inntrykk av: For Kim må relativisere sine reduksjoner ikke bare til bestemte systemer, men også til disse systemene ved bestemte tidspunkter. Det er nemlig et velkjent poeng innen multippel realiserbarhetsargumenter at hjernen er svært plastisk, og f.eks. kan kompensere for skader i ett område, ved å la andre områder overta det skadede områdets funksjon. Kim er taus om dette her.<sup>181</sup>

Kim aksepterer faktisk at psykologi på denne måten ikke blir en enhetlig vitenskap: For han hevder nettopp at ”the causal inheritance principle”, og multippel realiserbarhet impliserer at multippelt realiserbare egenskaper ikke er ”unified and robust properties in their own right”, men ”are sundered into their diverse realizers”. (MPW 111.) I Kim (1992) argumenterer han mer utførlig for dette, og slutter at psykologiens enhet bare er enheten av en konjunksjon av lokale psykologier. I utgangspunktet ville vi forvente at samme psykologi kunne gjelde for marsmenn, mennesker og kanskje for noen datamaskiner som kommer på markedet om 100 år. Men ifølge Kim vil den korrekte beskrivelsen være at vi har en Mars-psykologi & en menneske-psykologi & en datamaskin-psykologi.

Imidlertid ville Kim neppe akseptere at reduksjonene ikke bare er lokale, men også unike. Han vil formodentlig tvert imot mene at den lokale subveniente teorien er generell, og antagelig også en del av en mer global fysisk teori. Og til Kims forsvar kan man peke på at refleksjon over multippel realiserbarhet i forbindelse med kunstige gjenstander *ikke* støtter konklusjonen om at de lokale subveniente teoriene er unike: En isøks er f.eks. et objekt med en bestemt funksjon. (Den skal kort sagt kunne brukes til å rydde vekk, og til å skape oppheng i is.) Denne funksjonen er også multippelt realiserbar – det finnes økser med forskjellig form på skaftet, økser med forskjellig vinkel på og legeringer i øksebladet etc., etc. Men i hvert

---

<sup>180</sup> Emergentister vil selvsagt ikke godta Kims identiteter, men ideen om en unik lov kan kanskje brukes til å vise at hverken reduksjon via brolover eller via lokale identiteter lukker den epistemologiske kløften.

<sup>181</sup> Dvs. i Kim (1999), men se Kim (1992) s.329, der han prøver å besvare innvendingen.

enkelt tilfelle vil det utvilsomt finnes en nokså generell fysisk teori om hvordan øksene realiserer denne funksjonen. Det er ikke nødvendig å trekke inn unike lover for å forklare hvorfor en bestemt fysisk struktur oppfyller funksjonen vi krever av en isøks. Derfor finnes det heller ikke noen autonom, og enhetlig ikke-fysisk vitenskap om isøkser. Tvert imot er isøkser fysiske objekter, i den vide forstand av ordet fysisk som Kim forutsetter. "Isøks" er et begrep vi bruker for å plukke ut objekter som har fysiske egenskaper, som gjør at de kan brukes til visse formål. Det er ofte upraktisk å snakke om disse egenskapene (derfor har vi begrepet), men å være en isøks er ikke en egenskap forskjellig fra disse egenskapene: Et objekt har ingen kausale krefter i kraft av å være en isøks, som det ikke har i kraft av å ha visse fysiske egenskaper. Når det gjelder isøkser, kan Kim altså få det som han vil, med både lokale og forklarende reduksjoner.

Skal vi unnslippe Kims konklusjon, må vi derfor vise at emergente egenskaper ikke er i samme båt som de egenskapene ved hvilke vi identifiserer en kunstig gjenstand.<sup>182</sup> En måte å gjøre dette på ville være å vise at de er reelle egenskaper forskjellige fra baseegenskapene. Dette bringer meg til mitt angrep på Kims lokale reduksjonisme: Kims reduktive konklusjon følger fra hans funksjonalisme. For å unngå denne konklusjonen må emergentisten enten forkaste "the causal inheritance principle" (noe jeg ikke vil gjøre), eller å hevde at emergente egenskaper *ikke er annenordens funksjonelle egenskaper i Kims forstand*.

Det siste alternativet betyr altså at emergente egenskaper ikke kan defineres ut fra den kausale rollen til de mikrobaserte egenskapene de er superveniente på. Dette er egentlig ikke så overraskende gitt diskusjonen av emergensbegrepet ovenfor, men en viktig konsekvens av det er at *emergentisme ikke er forenelig med "annenordens" funksjonalisme*.<sup>183</sup> Dette er kanskje en pris enkelte ikke er villige til å betale, fordi mentale egenskaper dermed ikke gies den tette forbindelsen til fysiske egenskaper som funksjonalisme tilbyr. Emergentisme innebærer en sterkere form for *egenskapsdualisme* enn funksjonalisme. På den andre siden ser det ut til at dette er prisen man må betale dersom man ønsker å gi det mentale realitet og autonomi. For Kims argumenter synes klart å tilsi at dersom man vil være funksjonalist, så får man vær så god finne seg i å være lokal reduksjonist også.

Men selv om emergentisten må ha et sterkere dualistisk syn enn funksjonalisten, så trenger han heldigvis ikke benekte at mentale egenskaper må gies en (i hvert fall delvis)

---

<sup>182</sup> Det samme må funksjonalisten, hvis han vil unngå forklarende lokale reduksjoner av sine funksjonaliserte mentale egenskaper. Men denne utveien er neppe tilgjengelig for ham, gitt maskinanalogiers rolle i funksjonalisme.

<sup>183</sup> Kim konkluderer også med at emergens betyr nonfunksjonaliserbarhet. Kim (1999) s.18. Reduksjonsmodellen ovenfor er derfor bare rettet mot funksjonaliserbare egenskaper.

funksjonell rolle. Han kan nemlig definere en mental egenskap som egenskapen å ha en *emergent* egenskap som fyller en bestemt kausal rolle. I så fall er basemengden **B**, som denne annenordens egenskapen er *definert over*, forskjellig fra den basemengden som utgjør dens *supervenienbase*.<sup>184</sup> Egenskapen er følgelig bare annenordens i en triviell forstand. I det hele tatt må emergente egenskaper altså være *førsteordens* egenskaper essensielt.

På denne måten unngår emergentisten at det kan gies en lokal forklaring av hvorfor den mentale egenskapen instansieres. Men han unnslipper bare dette forsåvidt som den mentale egenskapen har nye kausale krefter. (Jf. den ontologiske komponenten.) Emergentisten kan ikke besvare kritikken av den epistemologiske komponenten, med mindre han kan besvare kritikken av den ontologiske komponenten: Emergentisme står og faller med DC.

## 7 Kims argumenter mot "downward causation"

### 7.1 Innledning

Kims argument mot emergensbegrepets ontologiske komponent kan sies å utgjøres av hans to argumenter mot DC, nemlig superveniensargumentet og det jeg kaller det asymmetribaserte argumentet. Disse argumentene er tema for dette kapittelet. DC er, som vi har sett, en *essensiell* forpliktelse for emergentisten. Det blir derfor viktig for ham å på et vis kunne unngå konklusjonen i argumentene mot DC. Underveis i kapittelet vil jeg forsøke å komme frem til hva som skal til for at han kan klare dette.

### 7.2 Kims superveniensargument mot "downward causation"

Jeg vil her se på Kims første og eksklusjonsbaserte argument mot DC. I MPW tar dette argumentet form av det Kim kaller "supervenienargumentet" (MPW 38-47.)<sup>185</sup> Argumentet skal vise at dersom mentale egenskaper er irreduisible, så er de epifenomenale. Jeg vil gi en oppsummering av dette argumentet med vekt på hvilke premisser som er virksomme.<sup>186</sup> Det

---

<sup>184</sup> Dette er en mulighet Kim tilsynelatende overser når han antyder at mange makroegenskaper som har blitt kalt emergente ikke fortjener tittelen, fordi det ikke er klart at de er nonfunktjonaliserbare. (Kim (1999) s.18) Men funksjonen må ikke nødvendigvis spesifiseres ut fra egenskapene de er superveniente på!

<sup>185</sup> Dette eksklusjonsbaserte argumentet finnes i forskjellige former i flere artikler og kan sies å utgjøre ryggraden i Kims angrep på DC. I Kim (1993d) formuleres argumentet ut fra *realisering*, og ikke superveniens, i Kim (1999) er det *emergens* som danner grunnlaget for argumentet.

<sup>186</sup> Jeg har – som vi etter hvert skal se – hatt stor nytte av Hansens klare diskusjon av hvilke premisser Kim bruker. Hansen (2000) s.470-473.

vil bli klart at prinsippet om det fysiske domenes kausale lukkethet spiller en sentral rolle for å få konklusjonen. Kort sagt sier prinsippet at en kausal kjede av fysiske hendelser ikke kan inneholde mentale hendelser, og at fysiske kausalforklaringer derfor ikke behøver henvise til mentale hendelser. (MPW 37 & 40.) Nedenfor vil jeg analysere og kritisere dette prinsippet. Jeg vil også kritisere en forutsetning Kim gjør om at mentale egenskaper alene må være tilstrekkelige årsaker. Posisjoner som forkaster disse to forutsetningene vil *ikke* berøres av superveniensargumentet.

Min kritikk av superveniensargumentet er altså direkte og metafysisk. Jeg diskuterer ikke det populære indirekte motargumentet som Kim kaller ”generaliseringsargumentet.” (MPW 80.) Dette går ut på at hvis superveniensargumentet viser at mentale egenskaper er epifenomenale, så vil også andre superveniente egenskaper som f.eks geologiske egenskaper være epifenomenale. Dette antas å være absurd, og man behøver derfor angivelig ikke bekymre seg over superveniensargumentet.<sup>187</sup>

Vi har allerede sett (5.4, MPW 41-43) første trinn i Kims superveniensargument: Det gikk ut på at årsakssammenhenger mellom egenskaper på samme nivå forutsetter ”downward causation” (DC). Dersom en mental egenskap M skal forårsake en annen mental egenskap M\*, så må dette altså skje ved å forårsake M\*’s superveniensbase P\*, dvs. ved DC. For å nå denne konklusjonen brukte Kim prinsippet om eksplanatorisk eksklusjon, som sier at det ikke kan finnes to komplette og uavhengige forklaringer av samme eksplanandum.

Resten av superveniensargumentet går ut på å vise at DC der M forårsaker P\* er umulig, fordi M’s egen superveniensbase P også er en årsak til P\*. Vi har dermed *to* antatte årsaker til P\*, og Kim argumenterer for at P *ekskluderer* (dvs. utelukker) M som årsak. Under forutsetning av at M ikke lar seg redusere til P (MPW 46), så er M dermed et epifenomen, dvs. kausalt impotent. Konklusjonen blir at den virkelige kausale prosessen foregår på subvenient nivå, og består i at P forårsaker P\*. (MPW 45.) Dermed vil DC fra mentale egenskaper være umulig. Dette vil være dårlige nyheter for emergentister.

Hvordan når Kim denne konklusjonen? Han argumenterer først for at ”both *M* and *P* seem severally eligible as a sufficient cause of *P*\*.” (MPW 43-44.) Han hevder at dette følger enten man forstår kausalitet som nomologisk tilstrekkelighet, eller ut fra kontrafaktiske påstander: P er nomologisk tilstrekkelig for M (dette er en egenskap ved superveniens), som

---

<sup>187</sup> Se Burge (1993) s.102, Fodor (1989) s.137-141, MPW 77-80 eller Van Gulick (1993) s.249 for varianter av argumentet. Kim vier stor plass til å besvare argumentet. (Spes. MPW 77-87 & 112-118.) Det er vel og bra dersom generaliseringsargumentet virker, men en direkte metafysisk kritikk av superveniensargumentets forutsetninger har den fordelen at den vil vise *hvorfor* argumentet (eventuelt) slår feil ut. Jeg er faktisk forberedt på å ta en hard linje med hensyn til generaliseringsargumentet: Hvis geologiske egenskaper har samme problemer som psykologiske egenskaper, så viser ikke dette annet enn at geologer også har sine problemer.

igjen (formodentlig ifølge antagelsen om at M er årsak til P\*) er nomologisk tilstrekkelig for P\*. Følgelig er P også nomologisk tilstrekkelig for P\*. Derfor må både P og M regnes som tilstrekkelige årsaker til P\*.<sup>188</sup> Det avgjørende her er modaliteten ved sterk superveniens; nemlig at P *nødvendigvis* er tilstrekkelig for M. Kim holder seg altså foreløpig trygt innenfor forutsetningene av minimal fysikalisme.

Nå går Kim over til å diskutere mulige relasjoner mellom de to årsakene M og P. En undertrykket forutsetning er altså at han er *berettiget* til å spørre om relasjonen mellom to årsaker til – og dermed mellom to kausalforklaringer av – samme hendelse. Selv om dette (kanskje beklageligvis) ikke går klart frem av selve superveniensargumentet, så tror jeg han her bygger på prinsippet om eksplanatorisk eksklusjon. For i et senere kapittel bemerker han i forbindelse med Tyler Burge at når to kausalforklaringer pretenderer å være komplette forklaringer av samme hendelse, har vi en "unstable situation requiring us to find an account of how the two purported causes are related to each other." (MPW 65.) Og han viser (ibid. n.12) til en artikkel hvor han formulerer dette prinsippet.<sup>189</sup> Prinsippet forplikter, som vi har sett, Kim til nå å vise at relasjonen mellom M og P er slik at de *ikke* gir komplette og uavhengige forklaringer av P\*. (Han kunne selvsagt ha gjort det lettere for seg selv, og eliminert M som forklaring.) Det som får denne delen av superveniensargumentet i gang, er altså et prinsipp som jeg har akseptert ovenfor, forsåvidt som prinsippet omhandler kausalforklaringer.

Kim diskuterer nå tre mulige *antireduktive* relasjoner mellom M og P.<sup>190</sup>

- (i) Vi har en kausal kjede fra P via M til P\*.
- (ii) M og P utgjør til sammen en tilstrekkelig årsak til P\*.
- (iii) P\* er *overbestemt* fordi M og P er tilstrekkelige hver for seg. (MPW 44.)

Han utelukker (i) ved å argumentere for at superveniensrelasjonen mellom P og M ikke er kausal. Derfor kan ikke P *forårsake* M. Dessuten vil M ifølge (i) være en nødvendig ikke-fysisk årsak til P\*, og dette bryter med det fysiske domenes kausale lukkethet. (Ibid.)

---

<sup>188</sup> Tilsvarende gir superveniens det mønster av nødvendige relasjoner mellom egenskapsinstansieringer som den *kontrafaktiske* analysen av kausalitet krever. Det er *kanskje* mulig å benekte at P konkurrerer med M som årsak, fordi P's bidrag i en eller annen forstand er *indirekte*: Vokteren som ga Sokrates giften, konkurrerer således ikke med giften om å forårsake Sokrates' død, men bidrar indirekte til denne hendelsen. Sokrates ville ikke ha dødd uten giften, og vokteren er en betingelse for giften, liksom P er en betingelse for M. Jeg vil imidlertid ikke forutsette dette i det følgende.

<sup>189</sup> Kim (1989a).

<sup>190</sup> Vi må huske at superveniensargumentet starter fra antagelsen om antireduksjonisme. (MPW 46.) Hvis derimot M=P, så oppstår ikke problemet om eksklusjon. Derfor er det overflødig å se på *reduktive* relasjoner mellom M og P i argumentet.



Relasjon (ii) er heller ikke mulig: For det første hevder han at M og P i følge (ii) er tilstrekkelige hver for seg. Han sier også at han ikke kan se hvordan de kan ha mer kausal kraft *sammen* enn de har hver for seg. Dessuten vil (ii) også bryte med den kausale lukketheten, siden M er en nødvendig del av årsaken. (Ibid.)

Mot (iii) innvender han at systematisk overbestemmelse i seg selv er implausibelt. Dessuten gjør (iii) M unnværlig ("dispensable") som årsak, fordi P ville kunne forårsake P\* alene. Kim har også et modalt argument for å vise at overbestemmelse står i konflikt med kausal lukkethet: For la oss se på en mulig verden der vi har M, men ikke P, men som ellers er så lik vår verden som mulig. I denne verden vil M forårsake P\* alene, og den fysiske lukketheten er brutt. Kim mener at vi ikke kan akseptere at "[...] a minimal counterfactual supposition like that can lead to a major change in the world." (Ibid.-MPW 45.) Jeg kan imidlertid ikke se at dette siste argumentet mot overbestemmelse holder: For siden M er supervenient på P, så er M nomologisk avhengig av P (eller en annen fysisk baseegenskap). M kan derfor *ikke* instansieres uavhengig av P (eller en annen fysisk baseegenskap) i verdener med samme fysiske lover som den faktiske verden. Å fjerne P er derfor ikke en "minimal" forandring, men en *stor* forandring: Kim har beveget seg utenfor verdener med samme naturlover som den faktiske verden, og dessuten forpliktet seg til eksistensen av en kartesiansk sjel. Gitt så store forandringer er det ikke overraskende at den kausale lukketheten brytes!<sup>191</sup> Men uavhengig av problemet om kausal lukkethet, så synes Kim å ha rett i at det er noe utilfredsstillende i at all "mental causation" er overbestemt: Hver *eneste* gang våre mentale tilstander forårsaker noe så er de overflødige, siden virkningen ville forekommet også uten den mentale tilstanden.

Etter å ha utelukket alternativene (i)-(iii) hevder Kim at relasjonen mellom M og P i stedet må være slik: (iv) "*P caused P\*, and M supervenes on P and M\* supervenes on P\*.*" (MPW 45.) En uheldig konsekvens av (iv) er at psykologiske forklaringer ikke blir *reelle*. For vi har stilt et strengt krav til psykologiens realitet; nemlig at den må referere til kausalt virksomme egenskaper. Ifølge (iv) er det P, og ikke M som gjør den kausale jobben og derfor er kausalt virksom. En psykologisk kausalforklaring som viser til M kan selvsagt *fungere* (f.eks. til forutsigelsesformål) fordi det er regulære og nødvendige sammenhenger mellom M og passende fysiske egenskaper som P. Men denne forklaringen vil mangle *realitet*. Som Kim

---

<sup>191</sup> Kim vil imidlertid kunne si at lukketheten brytes, til tross for at hans modale argument ikke holder: For i Kim (1989a) s.252 hevder han noe nølende at en forklaring av en overbestemt virkning ikke er komplett uten å nevne *begge* årsakene. En komplett forklaring av den overbestemte virkningen må følgelig nevne en ikke-fysisk årsak, og dermed vil Kim kunne hevde at lukketheten brytes.

sier, vil sammenhengene mellom M og de andre egenskapene være "parasitic on real causal processes".<sup>192</sup> (Ibid.) Dette er mildest talt utilfredsstillende.

### 7.3 Superveniensargumentets forutsetninger

Før jeg drøfter prinsippet om det fysiske domenes kausale lukkethet, må jeg se nærmere på hvor avgjørende dette og andre prinsipper er for Kim. La oss begynne med å avdekke betydningen til prinsippet om kausal lukkethet.

Dette prinsippet ble brukt en rekke ganger i superveniensargumentet. Likevel kan det se ut som om Kim mener han kan være det foruten. Det er i den forbindelse interessant å merke seg at prinsippet ikke har fått plass i indeksen i MPW. Hansen påpeker at: "In connection with the exclusion problem, at least, Kim seems inclined to downplay the significance of the causal closure principle."<sup>193</sup> ("The exclusion problem" (MPW 37-38 & 64-65) er i likhet med superveniensargumentet et eksklusjonsbasert argument. For våre formål kan det betraktes som en strukturelt lik kortversjon av superveniensargumentet.)

Hansen viser til Kims påstand om at alt som kreves for at vi skal ha et eksklusjonsproblem er at vi tror det finnes tilfeller der en virkning har, eller må ha, en mental og en fysisk årsak. (MPW 66.) Kim hevder videre at: "[...] the physical causal closure only makes the exclusion problem a general one for all mentalistic explanations of physical behaviour." (Ibid.) Kim må ha rett i at prinsippet gjør problemet *generelt*: For det garanterer at alle fysiske hendelser har en fysisk årsak. Derfor vil det være en fysisk årsak i *alle* tilfeller der vi antar at en fysisk hendelse har en mental årsak.

Men dette er ikke som Kim antyder *alt* prinsippet gjør: Vi har sett at han også bruker det i superveniensargumentet til å utelukke relasjoner mellom M og P der M forutsettes å spille en nødvendig rolle. Hansen hevder således at prinsippet er svært viktig for Kim: "[...] the principle of physical causal closure plays an essential role in the exclusion problem, just as it does in the supervenience argument."<sup>194</sup> Hansen argumenterer overbevisende for at begge av Kims formuleringer av eksklusjonsproblemet forutsetter prinsippet: (1) Kim bruker det "explicitly and ineliminably"<sup>195</sup> i sin første fremstilling av eksklusjonsproblemet (MPW 37-38.) (2) Men han trenger prinsippet (samt "the unacceptability of causal overdetermination")

---

<sup>192</sup> Jeg har tidligere stilt et spørsmål om hvordan vesensforskjellige vitenskaper som fysikk og psykologi kan fungere som teorier om den samme verden. Til en viss grad er (iv) et svar på dette spørsmålet. Men hvis (iv) er riktig, så er psykologi hverken autonom eller reell i den forstand jeg har forutsatt.

<sup>193</sup> Hansen (2000) s.471. (Se forøvrig ibid. s.461-463 for en utførlig diskusjon av eksklusjonsproblemet og dets forhold til superveniensargumentet.)

<sup>194</sup> Hansen (2000) s.472.

<sup>195</sup> Ibid. s.471.

også i sin andre fremstilling av problemet (MPW 64-65), selv om han ikke her bruker det eksplisitt. Dette skyldes at han forutsetter det implisitt, når han utelukker mulige antireduktive relasjoner mellom den fysiske og den mentale årsaken.<sup>196</sup>

Så prinsippet er essensielt i den forstand at konklusjonen ser ut til å følge hvis det aksepteres. En kritikk av argumentet må derfor enten avvise det, eller vise at antireduksjonisme likevel ikke bryter med det. På den andre siden er det ikke essensielt i den forstand at konklusjonen ikke følger *uten* det. For Kim har, som vi har sett, ett eller flere argumenter som er uavhengig av prinsippet, for hver mulige relasjon mellom M og P han utelukker.

Når det gjelder prinsippet om eksplanatorisk eksklusjon, er det klart at dette er nødvendig for Kim: Som vi husker sier dette prinsippet at: "[...] two or more complete and independent explanations of the same event or phenomenon cannot coexist."<sup>197</sup> Vi har sett at det som fikk problemet i gang, var nettopp at prinsippet forpliktet Kim til å finne en relasjon mellom M og P, slik at de ikke representerer komplette og uavhengige kausalforklaringer. Dersom de kunne eksistere side om side som komplette og uavhengige forklaringer, ville ikke P kunne ekskludere M, og psykologiske forklaringer ville være strengt autonome.<sup>198</sup>

Imidlertid tror jeg faktisk Kim mener at prinsippet om eksplanatorisk eksklusjon er *tilstrekkelig* for hans argument: Dette harmonerer med at han, som vi så, vektlegger at det er tilstrekkelig at vi har en mental *og* en fysisk årsak, for at problemet skal oppstå. (MPW 66.) Dessuten hevder han at man ikke trenger "much heavy-duty metaphysics" for å få eksklusjonsproblemet, men at det oppstår ganske enkelt fra *begrepet* om kausalforklaring og fra kausalrelasjonen:

[The exclusion problem] arises from the very notion of causal explanation and what strikes me as a perfectly intuitive and ordinary understanding of the causal relation. If this is right, turning away from metaphysics to embrace epistemology, or away from causation to embrace explanation, will not dissipate the need for an account of mental causation. (MPW 67.)

Det synes klart at det her er prinsippet om eksplanatorisk eksklusjon han tenker på. Ifølge dette prinsippet må alle som er villige til å innrømme at noe (f.eks. adferd) har både en fysisk og en mental forklaring, forholde seg til superveniensargumentet/eksklusjonsproblemet.

Det strategiske motivet for å vektlegge dette prinsippet fremfor det mer metafysiske prinsippet om kausal lukkethet synes like klart: Fordi problemet kan genereres fra et

---

<sup>196</sup> Ibid. s.472

<sup>197</sup> Kim (1989a) s.250.

<sup>198</sup> Kunne Kim klart seg med et svakere prinsipp, som likevel tilsa at det er en slags konflikt mellom M og P? I så fall ville dette bare styrke Kims posisjon. Men et svakere prinsipp kunne fort gitt antireduksjonisten mer spillerom.

eksplanatorisk prinsipp, slipper man ikke unna det ved f.eks. å hevde at eksplanatorisk praksis veier tyngre enn metafysiske bekymringer, slik Burge gjør.<sup>199</sup> Burge benekter på dette grunnlaget at mentale årsaker må *interferere* med fysiske kausale kjeder (jf. prinsippet om lukkethet), og at psykologiske forklaringer må *konkurrere* med fysiske forklaringer (jf. prinsippet om eksplanatorisk eksklusjon):

It would be perverse to think that such mental events must interfere with or alter, or fill some gap in, the chain of physiological events [...]. It would be perverse to think that the mentalistic explanation excludes or interferes with non-intentional explanations of the physical movement. [...] They [i.e. the allegedly *perverse* ideas] seem perverse because we know that the two causal explanations are explaining the same physical effect as the outcome of two very different patterns of events. The explanations of these patterns answer two very different types of enquiry. Neither type of explanation makes essential, specific assumptions about the other.<sup>200</sup>

Men hvis prinsippet om eksplanatorisk eksklusjon holder, så må Burge gå med på å diskutere relasjonen mellom den fysiske og den psykologiske forklaringen. Kim har altså sterke ambisjoner: Han ønsker at superveniensargumentet også skal berøre filosofer som er mindre ontologisk orienterte enn ham selv.<sup>201</sup>

Det er imidlertid ikke gitt at Kims forståelse av *kausalrelasjonen* nødvendigvis er så ”intuitive and ordinary” som han vil ha det til. Hansen hevder plausibelt at Kims henvisning til intuisjoner om kausalitet er problematisk.<sup>202</sup> Og gitt at våre intuisjoner om kausalitet formodentlig *stammer fra* våre forklaringspraksiser, vil praksis-orienterte filosofer kunne prøve å innvende følgende: I den grad Kims forståelse av kausalitet kommer i konflikt med praksisene, så må Kims forståelse revurderes. Jeg vil ikke forutsette noen slik kritikk, men vil i forbindelse med Sperry (8.4) komme med noen få bemerkninger om kausalitet.

Hva med premisset om at systematisk overbestemthet er uakseptabelt? Dette premisset ligger åpenbart i kortene når Kim utelukker systematisk overbestemmelse som en mulig relasjon mellom M og P. (MPW 44-45.) Men som vi så, utelukket Kim overbestemmelse *delvis* fordi han mente at dette ville innebære et brudd på prinsippet om kausal lukkethet. (MPW 45.) I en viss forstand er prinsippet derfor av sekundær betydning i forhold til prinsippet om lukkethet. Likevel er det klart at dets fravær ville svekket argumentet. Det som ikke er klart er i hvilken grad Kim er avhengig av å hevde at systematisk overbestemmelse er *metafysisk* uakseptabelt (i seg selv, eller pga. et eventuelt brudd med

---

<sup>199</sup> Kim kritiserer Burge for dette. (MPW 65-67.) Burge synes å hevde at vi har større grunn til å tro på våre eksplanatoriske praksiser enn på det materialistisk metafysikk forteller oss. (Burge (1993) s.103.)

<sup>200</sup> Burge (1993) s.115-116.

<sup>201</sup> En annen sak er at Burge faktisk ser ut til å være åpen for muligheten av en *slags* metafysikk om relasjonen mellom forklaringer: Han er imidlertid skeptisk til at relasjonen mellom mentale og fysiske hendelser kan være identitet eller den mereologiske relasjonen. (Ibid. s.116.) Men dermed setter han seg – sikkert bevisst – på sidelinjen av fysikalistiske diskusjoner.

<sup>202</sup> Hansen (2000) s.470.

kausal lukkethet). Han kan antagelig hevde at det at overbestemmelse gjør M unnværlig som årsak, rett og slett er *eksplanatorisk* utilfredstillende eller problematisk.

Den helt sentrale rolle Kim gir prinsippet om eksplanatorisk eksklusjon er av en viss interesse: Dette skyldes at mange diskusjoner av eksklusjonsproblematikk fokuserer på metafysiske prinsipper om kausal lukkethet og overbestemmelse, mens Kim altså toner ned disse prinsippene, slik at hans argument ikke er strengt avhengig av dem.

Kim har etter min mening argumentert plausibelt for at P truer med å ekskludere M som årsak. Under forutsetning av *prinsippene om eksplanatorisk eksklusjon og kausal lukkethet* ser det ut til å følge at M er epifenomenal. Jeg aksepterer det første prinsippet (for kausalforklaringer), men vil nedenfor betvile om minimal fysikalisme må akseptere det siste.

En positiv historie om hvordan DC foregår, vil imidlertid også måtte finne et smutthull i argumentasjonen som er uavhengig av prinsippet om kausal lukkethet: I utgangspunktet synes det for meg mest lovende å bygge videre på alternativ (ii) der M og P til sammen utgjør en tilstrekkelig årsak. Jeg vil kalle dette alternativet for ”arbeidsdelingsmodellen for DC.” Kim utelukket denne modellen med henvisning til kausal lukkethet, men *også* fordi M og P hver for seg er tilstrekkelige. Hansen påpeker i sin oppsummering av argumentet at dette formodentlig er ”*ex hypotesi*”.<sup>203</sup> At M skal være en tilstrekkelig årsak uavhengig av P, er dermed noe Kim har bakt inn i forutsetningene for argumentet.<sup>204</sup> Men er det ikke mulig å revurdere denne forutsetningen, og innrømme at hvis M skal forårsake noe, så må dette være i en forstand avhengig av, og sammen med superveniensbasen P? Kort sagt; er det ikke mulig at det er en arbeidsdeling mellom M og P?

Eksklusjonsproblemet oppstår selvsagt likevel dersom P *alene* er tilstrekkelig. Men Kims argument (ovenfor, MPW 43) for at P alene er en tilstrekkelig årsak, bygger på antagelsen om at M alene er en tilstrekkelig årsak. Og denne antagelsen vil jeg altså – i det minste forsøksvis – revurdere. Kanskje er det en slik mulighet Hansen hentyder til i forbindelse med første trinn i superveniensargumentet når han sier at ”[...] the idea of a sufficient cause surely needs to be subjected to scrutiny anyway.”<sup>205</sup>

Til slutt vil jeg bemerke at Sperrys og Van Gulicks modeller for DC (8.2-8.5) antagelig innebærer at M og P nettopp *til sammen* utgjør en tilstrekkelig årsak.

---

<sup>203</sup> Ibid. s.468.

<sup>204</sup> Det holder ikke å innvende at antagelsen er bakt inn i *begrepet om årsak*. For poenget med arbeidsdelingsmodellen er ikke at det ikke *finnes* en tilstrekkelig årsak, men at hverken M eller P *alene* er en slik årsak.

<sup>205</sup> Hansen (2000) s. 467.

## 7.4 Prinsippet om det fysiske domenes kausale lukkethet

Jeg vil her se på Kims begrunnelse for at alle fysikalister er forpliktet til at det fysiske domene er kausalt lukket. Deretter argumenterer jeg for at dette prinsippet om kausal lukkethet er sterkere enn minimal fysikalisme, og at det finnes grunner til å betvile det. Ovenfor så vi at store deler av superveniensargumentet er avhengig av dette prinsippet. Hvis min kritikk av prinsippet er tilfredsstillende, vil mye av argumentets brodd forsvinne: For DC vil ikke stå i noen motsetning til den *fysikalistiske* delen av minimal fysikalisme.

Kim gir følgende formulering av "the principle of physical causal closure":

If you pick any physical event and trace out its causal ancestry or posterity, that will never take you outside the physical domain. That is, no causal chain will ever cross the boundary between the physical and the nonphysical. (MPW 40.)

Kims standardbegrunnelse for prinsippet er at hvis man benekter det, så må man også benekte "the in-principle completability of physics", hvilket er uakseptabelt for en "serious physicalist". (MPW 40.)<sup>206</sup> Av dette må det følge at *minimal* fysikalisme er forpliktet til prinsippet. (Spesielt siden Kims argument er rettet mot antireduksjonister som mener at minimal fysikalisme er tilstrekkelig.)

La oss se nærmere på begrunnelsen for prinsippet, som kommer særlig klart frem i Kim (1993d). Her starter han fra det utvilsomt riktige premisset om at et brudd på prinsippet innebærer en kartesiansk kausal dualisme, forstått som at samme kausale kjede kan inneholde både fysiske og mentale hendelser.<sup>207</sup> Men fra dette slutter han:

It follows then that under Cartesian causal dualism there can be *no complete physical theory of physical phenomena*. For it allows physical occurrences that cannot be causally explained by invoking physical antecedents and laws alone. Any comprehensive theory of the physical world must, on Cartesian interactionism, include references to non-physical causal agents and laws governing their behaviour. We can say then that *Cartesian interactionism violates the causal closure of the physical domain*.<sup>208</sup>

Hvilket begrep om "complete" og "comprehensive" fysisk teori benytter Kim seg av her?<sup>209</sup> For det første er hans begrep åpenbart forenelig med minimal fysikalisme i den forstand at muligheten for ikke-fysiske forklaringer (dvs. teoretisk pluralisme) fortsatt står åpen: Kim krever ikke at alt skal kunne gies en fysisk forklaring, men at alt *fysisk* skal kunne gies en

---

<sup>206</sup> Samme begrunnelse finnes i f.eks. Kim (1989b) s.280 eller (1998b) s.147.

<sup>207</sup> Kim (1993d) s.188, se også MPW 37.

<sup>208</sup> Kim (1993d) s.189-190. Senere bemerker Kim at de fleste emergentister gladelig vil akseptere brudd på prinsippet, men at han tviler på om denne muligheten står åpen for "contemporary non-reductive physicalists". (Ibid. s.209.) Implisitt her er formodentlig en antagelse om at fullblods emergentister er forpliktet til en sterk form for egenskapsdualisme, og derfor ikke vil akseptere prinsippet.

<sup>209</sup> Termene "complete" og "comprehensive" ("komplett" og "omfattende") brukes også i MPW 40, så bruken av dem er neppe aksidensiell, men representerer tvert imot hva Kim mener vi prinsipielt kan forvente oss av fysikk.

fysisk forklaring.<sup>210</sup> Men for det andre krever han (mer problematisk) at forklaringer av *fysiske* hendelser ikke må henvise til *ikke-fysiske årsaker og lover*, de kan forklares ut fra ”physical antecedents and laws alone”.

Tanken bak prinsippet er altså at det fysiske domene utgjør et lukket og selv-tilstrekkelig (subvenient) system som ikke er avhengig av kausal ”input” fra ikke-fysiske (superveniente) egenskaper. Fysiske kausalforklaringer er dermed autonome i en svært streng forstand: De behøver ikke henvise til annet enn fysiske egenskaper.<sup>211</sup>

Bak dette ligger tilsynelatende en ontologisk idé om noe grunnleggende værende med røtter i antikken: Det fysiske avhenger ikke for sin væren *eller for sin forklaring* av det ikke-fysiske, mens det ikke-fysiske *er* avhengig av det fysiske på denne måten. Bildet av det fysiske som selv-tilstrekkelig i denne forstanden har interessante likheter med forholdet mellom platoniske ideer og sanseting.

La oss – i det minste for argumentets skyld – følge Kim i antagelsen om at fysikk må være prinsipielt kompletterbar. Anta videre at prinsippet om det fysiskes kausale lukkethet er brutt. Vi kan likevel tenke oss at vi har en teori T som inneholder *alle* lover der fysiske egenskapsinstansieringer inngår. Noen av disse lovene vil postulere årsakssammenhenger fra det ikke-fysiske til det fysiske (kausal ”input” til fysikk), andre vil postulere årsakssammenhenger motsatt vei (kausal ”output” fra fysikk). Men i dette lovsystemet vil det også være en kjernet teori T\*, som forteller oss alt som er å vite om relasjoner internt mellom fysiske egenskapsinstansieringer. T\* vil inneholde gode, gamle rent fysiske lover som ”kraft = motkraft” ved siden av en rekke nye rent fysiske lover. Det er ikke åpenbart for meg at vi som minimale fysikalister skal ha et begrep om ”komplett og omfattende fysikk” som er så sterkt som Kims. Tvert imot er T\* en komplett teori i den forstand at den gir ”formen” til, eller de generelle betingelsene for fysiske egenskapsinstansieringer. ”Innholdet,” eller kausal ”input” til T\*, kommer imidlertid delvis fra ikke-fysiske egenskapsinstansieringer. (Derfor er T\* kanskje bare ”komplett” og ikke ”omfattende” i Kims betydning.)

Davidson ser imidlertid ut til å begrunne prinsippet ut fra de pragmatiske hensyn fysisk vitenskap er uttrykk for: Davidson mener at fysikk er knyttet til ønsket om å finne maksimalt generelle og strenge lovmessige sammenhenger mellom hendelser. Følgende sitat fra ”Mental Events” tyder på at prinsippet om det fysiskes lukkethet nærmest er en betingelse for at fysikk skal kunne svare til denne interessen:

---

<sup>210</sup> Muligheten for teoretisk pluralisme ligger på samme måte i kortene i MPW 40.

<sup>211</sup> Merk at vi ikke er ute etter en så sterk autonomi for psykologi: Emergentisme er forenelig med at noen psykologiske forklaringer nevner fysiske hendelser.

[...] a law can hope to be precise, explicit, and as exceptionless as possible *only if* it draws its concepts from a comprehensive *closed* theory. This ideal theory may or may not be deterministic, but it is if any true theory is. Within the physical sciences we do find homonomic generalizations, generalizations such that if the evidence supports them, we then have reason to believe they may be sharpened indefinitely by drawing upon further *physical* concepts: there is a theoretical asymptote of perfect coherence with all the evidence, perfect predictability (*under the terms of the system*), total explanation (*again under the terms of the system*).<sup>212</sup>

For Davidson er lukkethet altså innebygget i det fysiske vokabularets natur. Det er ikke slik at man etter en tids fysisk forskning *oppdager* at det systemet vi beskriver med vokabularet er lukket, fordi det er lukket så å si per definisjon: Innenfor Davidsons filosofiske system er prinsippet om lukkethet et regulativt ideal, som fysisk praksis forutsetter.

Men vi kan følge Davidson i at fysikk er en vitenskapelig praksis som arbeider mot strenge og generelle sammenhenger, uten å anta absolutt lukkethet. For når vi bedriver fysikk er vi typisk sett ute etter slike lovmessige sammenhenger som lar oss forutsi et fysisk systems tilstand *gitt* dets fysiske tilstand ved et tidligere tidspunkt. (Eller vi er i ettertid interessert i å *forklare* den siste tilstanden ut fra den første.) Må vi dermed også hevde at den første tilstanden ikke kan ha en ikke-fysisk årsak?<sup>213</sup> Davidsons argumentasjon i "Mental Events" er ikke tilstrekkelig til å vise at dette er nødvendig for at *rent* fysiske lover skal kunne være strenge.

En annen innvending av pragmatisk natur vil være å si at vi ikke begynner å lete etter ikke-fysiske årsaker, når vi observerer en fysisk forandring. Det er rett og slett mest praktisk å søke årsaken i det fysiske domene, i det minste i første omgang. Dette høres rimelig ut, men det avhenger av konteksten for undersøkelsen: Hvis vi f.eks. har grunn til å tro at mentale hendelser kan ha påvirket fysiske hendelser, har vi ingen problemer med å se etter en ikke-fysisk årsak. Vi får et uforklarlig resultat i et eksperiment, men innser plutselig hvorfor: Forskeren har – med en intendert kroppslig bevegelse – forstyrret de fysiske kausale kjedene! Dette innebærer ikke at de rent fysiske lovene ikke kan være strenge. (Vi skal imidlertid se at mentale årsaker, i følge Sperry, kun forstyrrer fysiske kausalkjeder i en *kvalifisert* forstand av "forstyrre.")<sup>(8.2)</sup><sup>214</sup>

Også uavhengig av disse resonnementene kan vi se at Kims prinsipp er svært sterkt: Sammen med antagelsen om at egenskaper må ha kausale krefter for å være reelle, utelukker prinsippet nærmest *a priori* egenskapsdualisme. Det blir i hvert fall vanskelig å se hvordan mentale egenskaper skal kunne forårsake fysiske egenskaper ved DC. Detaljer må selvsagt

---

<sup>212</sup> Davidson (1970) s.219, mine kursiveringer.

<sup>213</sup> Davidson vil selvsagt være skeptisk til å snakke om at en hendelse forårsaker noe *qua* fysisk eller ikke-fysisk. (3.2)

<sup>214</sup> Det vil imidlertid fastholdes at de mentale egenskapene er *nødvendige* for at de fysiske kausalkjedene tar den veien de tar. Det er snakk om en arbeidsdeling mellom M og P, der M er en *nødvendig* del.



fylles ut – slik Kim gjør i superveniensargumentet – for å vise at dette virkelig er et problem. Likevel ser situasjonen for autonomi og antireduksjonisme med hensyn til mentale egenskaper svært alvorlig ut, allerede i det øyeblikk prinsippet aksepteres, før argumentasjonen er begynt.<sup>215</sup> Minimal fysikalisme skal i *utgangspunktet* være forenelig med antireduksjonisters forutsetninger, den skal være en felles metafysisk plattform man kan diskutere ut fra. Derfor ser jeg ikke at den minimale fysikalist må akseptere prinsippet om det fysiske kausale lukkethet uten at flere, (kanskje empiriske) argumenter er gitt.<sup>216</sup>

Men kan man ikke si at prinsippet følger fra den asymmetriske avhengighetsrelasjonen som minimal fysikalisme forutsetter? For superveniens tilsier nettopp at fysiske egenskaper *ikke* avhenger for sin væren av mentale egenskaper, mens mentale egenskaper *er* avhengige av fysiske egenskaper på denne måten. Vil ikke det å postulere at en ikke-fysisk egenskap M er årsak til en fysisk egenskap P gjøre det fysiske avhengig av det mentale? Nei, for superveniensrelasjonen omhandler en konstitutiv eller mereologisk avhengighet, den er taus om kausal avhengighet. P kan meget vel forårsakes av M uten å være konstitutivt-mereologisk avhengig av M (hvilket ville være absurd). Fra at en årsak forårsaker en virkning, følger ikke at virkningen ikke kunne ha eksistert uten årsaken: For anta at prinsippet om kausal lukkethet brytes ved at en mental egenskap M forårsaker en fysisk egenskap P, uten at P også har en fysisk årsak. Det følger ikke fra dette at P aldri kunne ha blitt instansiert uten en mental årsak, for i en annen mulig verden kunne P ha blitt forårsaket av en fysisk årsak P\*. Dermed er det ikke slik at P nødvendigvis er avhengig av mentale egenskaper for sin eksistens.

Den kausale uavhengigheten som prinsippet om kausal lukkethet gir det fysiske er altså sterkere enn den asymmetriske avhengighetsrelasjonen som ligger i sterk superveniens.<sup>217</sup> Supervenienstenes og prinsippet om lukkethet er konseptuelt og logisk sett distinkte påstander. Det er heller ikke slik at en begrunnelse av det ene nødvendigvis er en begrunnelse av det andre: Intuisjonene og resonnementene vi appellerte til i forbindelse med supervenienstenes involverte ikke noen *kausal* uavhengighet ved det fysiske, og impliserer heller ikke noen slik uavhengighet.

---

<sup>215</sup> For funksjonalisme ser ikke disse vanskelighetene så store ut: For her er mentale egenskaper annenordens egenskaper, og det er strengt tatt de fysiske baseegenskapene som har de virkelige kausale kreftene. Men som vi har sett (6.4), er det tvilsomt om funksjonalisme er metafysisk sterk nok til å sikre psykologien autonomi.

<sup>216</sup> Kim hevder også at det å ikke akseptere prinsippet vil "put the layered view of the world itself in jeopardy; it is likely to require some serious rethinking." (Kim (1993d) s.209-210) Jeg har ikke helt klart for meg hva slags vanskeligheter Kim tenker på her. Imidlertid vil jeg i forbindelse med hans andre argument mot DC behandle forholdet mellom DC og den mereologiske relasjonen som verden er ordnet etter.

<sup>217</sup> Men som vi skal se, benytter Kims andre argument mot DC seg av den avhengighetsrelasjonen som ligger i sterk superveniens, og bildet av en lagdelt verden.

Et annet mulig problem for prinsippet er at Kim beskylder mange antireduksjonister for å ha et for snevert syn på hva som er fysisk. I følge Kim innebærer det fysiske domene nemlig ikke bare mikrofysiske objekter, men også fysiske objekter i videre forstand:

Plainly the physical domain must also include aggregates of basic particles, aggregates of these aggregates, and so on, without end; atoms, molecules, cells, tables, planets, computers, biological organisms, and all the rest must be, without question, part of the physical domain.” (MPW 113.)

Problemet er ikke at disse objektene ikke er fysiske, men at Kims svar på generaliseringsargumentet forutsetter at fysikk i denne *vide* forstanden er kausalt lukket. (MPW 116-117.) Scott Sturgeon hevder imidlertid at det kan være problematisk å betrakte fysikk i vid forstand, og ikke bare mikrofysikk, som kausalt lukket: Man kan (ifølge Sturgeon) med støtte i vitenskapen, si at ”Every quantum effect has a fully revealing, purely quantum history.”<sup>218</sup> Imidlertid kan man ikke uten videre si at: ”Every broadly physical effect has a fully revealing, purely broadly physical history.”<sup>219</sup> Det er dette siste Kim hevder, og må hevde. Og her synes det å være rom for ytterligere debatt. Som Sturgeon sier:

This claim is *not* part of extant science; nor is it part of everyday experience. [...] both macro science and everyday experience rely upon *mental* causes for broadly physical effects. If such effects always have broadly physical causes, as this interpretation of ”physical” demands, an argument is needed to show they do[.]<sup>220</sup>

Spørsmålet om lokalisering av lukketheten viser at Kims prinsipp ikke nødvendigvis har den uproblematiske statusen han synes å tillegge det: Antagelig henter prinsippet den intuitive appell det måtte ha fra mikrofysikk, og ikke fra fysikk i vid forstand. Kim ser altså ut til å være avhengig av en sterkere og – noen vil mene – mer problematisk påstand enn påstanden om lukkethet på mikrofysisk nivå.

Alt i alt mener jeg å ha begrunnet påstanden om at Kims prinsipp er betydelig sterkere enn minimal fysikalisme. Ved å anta prinsippet har Kim derfor bakt svært mye ekstra inn i forutsetningene for superveniensargumentet. Og det er i beste fall villedende av Kim å først definere minimal fysikalisme som sterk superveniens (MPW 15), for så senere å innføre prinsippet som en tilleggsantagelse alle fysikalister er forpliktet til.<sup>221</sup> (MPW 37 & 40.)

---

<sup>218</sup> Sturgeon (1998) s.415.

<sup>219</sup> Ibid. s.416.

<sup>220</sup> Ibid. For å konstruere et best mulig eksklusjons/overbestemmelsesargument trenger man derfor et argument for at mentale årsaker har kvantemekaniske virkninger, eller for at fysikk i vid forstand er kausalt lukket. (Ibid.) Jeg kan ikke her gå nærmere inn på Sturgeons subtile diskusjon og kritikk av eksklusjons/overbestemmelsesargumenter. Et annet interessant aspekt ved spørsmålet om lokalisering av lukketheten finner vi hos Hansen. Han argumenterer for at Kims svar på generaliseringsargumentet forutsetter at *mikrofysikk* ikke er kausalt lukket, og hevder at det er her vi har best grunner for å tro på lukkethet. (Hansen (2000) s.488.)

<sup>221</sup> Dette kan kanskje til en viss grad forklares (men neppe unnskyldes) ved at han, som vi har sett, antagelig tror at prinsippet om eksplanatorisk eksklusjon er tilstrekkelig for argumentet. Ved MPW 119 synes han dessuten *indirekte* å åpne for at det kan finnes en slags annenrangs fysikalisme uten prinsippet om lukkethet. (For han

## 7.5 Prinsippets implikasjoner

Dersom jeg har rett angående prinsippets styrke og implikasjoner for mentale egenskapers kausale krefter, så ville man kanskje forvente at de fleste antireduksjonister ikke aksepterte det. Slik er det ikke. Davidson ser f.eks. ut til å akseptere prinsippet uten bekymring.<sup>222</sup> Dette til tross for at han kun er villig til å akseptere svak, og ikke sterk superveniens. Dvs. hans syn på forholdet mellom det mentale og det fysiske er på dette området svakere enn minimal fysikalisme. (Dette henger selvsagt sammen med hans syn på hendelser og kausalrelasjoner mellom hendelser, som jeg kritiserer i (3.2). Det ser faktisk ut til å herske nokså bred enighet om prinsippet blant antireduksjonister, til tross for at det meg bekjent ikke er gitt veldig omfattende begrunnelser for det. Det å benekte prinsippet ser ut til å være radikalt og kontroversielt. Det er også noe utilfredsstillende i å møte Kims argument på denne måten, fordi Kim neppe vil gi slipp på prinsippet.

Imidlertid ser det for meg ut som om prinsippet i det minste fortjener mer eksplisitt oppmerksomhet: Kims argumenter synes å vise at prisen antireduksjonister må betale for å akseptere prinsippet er epifenomenalisme. Prinsippet nyter – som Kim viser – en sentral rolle i et moderne bevissthetsfilosofisk dilemma, som formuleres elegant av Victor Caston:

Nonreductive materialists appear to be caught in a dilemma. Either mental events cannot bring about anything in virtue of being mental, a position that resembles epiphenomenalism; or they can, in which case we seem to have the emergence of new nonphysical causal powers. But embracing epiphenomenalism has always seemed anathema, except to the hard-hearted. The question is whether nonreductive materialists ought to feel any more comfortable in the arms of emergentists.<sup>223</sup>

Det ser altså ut som om man har følgende opsjoner med hensyn til prinsippet om det fysiske lukkethet: Enten er man reduksjonist og kan akseptere prinsippet uten bekymring.<sup>224</sup> Eller så er man antireduksjonist og har valget mellom: (i) Å akseptere prinsippet og betale med epifenomenalisme. (ii) Å fornekte prinsippet og få ”mental causation,” mot prisen av emergentisme: Dvs. antireduksjonisten må gå inn for en robust form for egenskapsdualisme, der mentale egenskaper er irreduisible og strengt ikke-fysiske, dersom han skal kunne ha

---

snakker om å være fysikalist ”at least to the extent of respecting the physical causal closure”). Det ville altså i vært høyst ønskelig at han tematiserte i hvilken grad han mener lukkethet er en nødvendig betingelse for fysikalisme.

<sup>222</sup> Se diskusjonen ovenfor, og Davidson (1970) s.219. Faktisk ser Burge også ut til å akseptere prinsippet. (Burge (1993) s.102.)

<sup>223</sup> Caston (1997) s.309-310. Se også Pihlström (kommende i 2002). Pihlström viser også til en nylig diskusjon av dilemmaet hos Stephan (1999) s.198-301. Dilemmaet går ifølge Pihlström/Stephan så langt tilbake som til Pepper (1926).

<sup>224</sup> Men da er prinsippet også trivielt: For dersom alle egenskaper er fysiske egenskaper, er det klart at de bare kan ha fysiske egenskaper som sine årsaker.

”mental causation.” Gitt dette tilsynelatende dilemmaet må det i det minste kunne sies å være behov for en mer omfattende debatt omkring prinsippet og dets begrunnelse.

## **7.6 Konsekvenser av analysen av superveniensargumentet**

Jeg har argumentert for at man kan være fysikalist uten å akseptere prinsippet om det fysiskes kausale lukkethet. Dermed kan ikke superveniensargumentet sies å vise at DC er uforenelig med minimal fysikalisme. Men de delene av argumentet som er uavhengig av prinsippet, legger føringer på hva slags modell for DC som er akseptabel: En måte å ta hensyn til disse føringene på er å revurdere forutsetningen om at mentale årsaker er tilstrekkelige årsaker, og bygge en arbeidsdelings modell for DC.

## **7.7 Kims asymmetribaserte argument mot ”downward causation”**

Meningen med Kims superveniensargument i MPW var å vise at DC fra irreducible mentale egenskaper til fysiske egenskaper er umulig. I de nye og tildels overlappende artiklene ”Making Sense of Emergence” og ”Making Sense of Downward Causation”, har Kim formulert et annet argument mot DC fra (emergente) mentale egenskaper.<sup>225</sup>

Argumentet kan sies å ta utgangspunkt i en intuisjon om at ideen om DC fra emergente egenskaper til baseegenskapene de er superveniente på, er sirkulær eller inkohherent:

Is it coherent to suppose that the presence of *X* is entirely responsible for the occurrence of *Y* (so *Y*'s very existence is totally dependent on *X*) and yet *Y* somehow manages to exercise a causal influence on *X*?<sup>226</sup>

Kim vil undersøke om slike intuitive argumenter har virkelig kraft.<sup>227</sup> Jeg vil påstå at kjernen i argumentet som følger består i å appellere til den asymmetriske avhengighetsrelasjonen minimal fysikalisme forutsetter: Makroegenskaper er avhengige for sin væren av mikrobaserte egenskaper, mens det omvendte ikke er tilfellet. Emergentister forutsetter bildet av en lagdelt verden fra mikro til makro, der den asymmetriske avhengighetsrelasjonen er innebygget. Argumentet benytter seg altså av essensielle metafysiske forutsetninger som emergentister må gjøre.

Første trinn i argumentet består i at Kim skiller mellom en form for DC som han mener er uproblematisk, og en han mener er problematisk. Deretter tilskriver han den

---

<sup>225</sup> Kim (1999) og Kim (2000).

<sup>226</sup> Kim (1999) s.25, se også s.28. Stephan finner det samme problemet i forbindelse med emergentisten Sperry. (Stephan (1992) s.45.)

<sup>227</sup> Kim (1999) s.25

problematiske varianten til den moderne nevrofysiologen og emergentisten Sperry: Han mener at Sperry er forpliktet til *refleksiv* (dvs. selv-rettet) DC, der hendelser ved et hele forårsaker hendelser ved heleets *egen* mikrostruktur.<sup>228</sup> *Nonrefleksiv* DC innebærer derimot kausal påvirkning på "lower-level entities that are not among its [dvs. heleets] constituents."<sup>229</sup> Hvis Kim f.eks. kaster en vase ut av vinduet, vil den gå i stykker mot fortauet og dessuten forstyrre luftmolekyler på vei ned.<sup>230</sup> Som Kim sier er disse virkningene – som skjer i kraft av at vasen veier 1 kg – lavere nivå relativt til egenskapen å veie 1 kg. Kim har ingen innsigelser mot denne formen for DC fra (resultante) makroegenskaper. (Jeg har hittil noe unøyaktig antydnet at Kim er imot DC *per se*, men jeg vil heretter skille mellom problematiske og uproblematiske former for DC.) I kraft av å ha en masse på 1 kg har vasen nemlig : "[...] a set of causal powers that none of its micro-constituents have; the causal powers that this property represents cannot be reduced to the causal powers of micro-constituents of its bearers."<sup>231</sup>

Det Kim synes å ville ha frem er at hans teori er forenelig med at makroobjekter har kausale krefter delene ikke har. I MPW sier han tilsvarende: "Physicalism need not be, and should not be, identified with micro-physicalism." (MPW 117) Det ser ut til at han i "Making Sense of Emergence" forsøker å legitimere denne påstanden fra MPW, ved å legitimere visse former for DC.

Imidlertid tillegger Kim som nevnt Sperry en tese om *refleksiv* DC fra emergente egenskaper, der de emergente egenskapene dessuten er *avhengige* av mikrostrukturen de kausalpåvirker. Kims bilde av Sperrys teori om DC blir dermed slik: W er et hele, der delene  $a_1, \dots, a_n$  har egenskapene  $P_1, \dots, P_n$ , og står i relasjonen R. M er en emergent egenskap ved W.

- (i) [Downward causation]  $W$ 's having property  $M$  causes some  $a_j$  to have  $P_j$ ; but
- (ii) [Upward determination] each  $a_i$ 's having  $P_i$  and  $R$  holding for the  $a_i$ s together determine  $W$  to have  $M$  - that is  $W$ 's having  $M$  depends wholly on (or is wholly constituted by) the  $a_i$ s having the  $P_i$  respectively and being related by  $R$ .<sup>232</sup>

Neste trinn i argumentet skal vise at (i) og (ii) er inkohrente. For å få denne konklusjonen skiller Kim ytterligere mellom *synkron* og *diakron* refleksiv DC. Ved synkron refleksiv DC skjer DC og "upward determination" *samtidig*, ved tiden  $t$ : Det at W har M ved  $t$ , forårsaker at  $a_j$  har  $P_j$  ved  $t$ . Men M er også avhengig av  $P_j$  ved  $t$ .<sup>233</sup> Ved diakron refleksiv

---

<sup>228</sup> Ibid. s.26.

<sup>229</sup> Ibid. s.27.

<sup>230</sup> Ibid. s.26.

<sup>231</sup> Ibid. s.26. Når Kim sier at de kausale kreftene til vasen er *irreducible*, må han mene at ingen av vasens deler har disse kausale kreftene *alene*. For fra delenes masser kan vi ved addisjon dedusere makroegenskapen å ha en masse på 1 kg. Dette vil også være en *forklarende* reduksjon av makroegenskapen.

<sup>232</sup> Kim (1999) s.27 Merk at Kim ser bort fra muligheten av DC der relasjonen R er det som kausalpåvirkes.

<sup>233</sup> Kim (1999) s.28

DC forårsaker derimot det at  $W$  har  $M$  ved  $t$ , at  $a_j$  har en egenskap  $Q$  ved et *senere* tidspunkt  $t+\Delta t$ .<sup>234</sup>

I forbindelse med synkron refleksiv DC kommer Kims intuisjon om inkohrens eller sirkularitet inn i bildet igjen: Han snakker om en følelse av ”metaphysical discomfort” når  $W$  ikke kunne hatt  $M$  uten  $P_j$  (dvs. ”upward determination”), og  $M$  samtidig skal forårsake  $P_j$  (dvs. DC).<sup>235</sup> Kim skjerper denne intuisjonen til et eksplisitt prinsipp, ”the causal-power actuality principle”:

For an object,  $x$ , to exercise, at time  $t$ , the causal/determinative powers it has in virtue of having property  $P$ ,  $x$  must *already* possess  $P$  at  $t$ . When  $x$  is caused to acquire  $P$  at  $t$ , it does not already possess  $P$  at  $t$  and is not capable of exercising the causal/determinative powers inherent in  $P$ .<sup>236</sup>

Prinsippet omhandler altså tilfeller der  $x$ 's kausale krefter til å gjøre noe er avhengig av at  $x$  (eller som vi skal se; en del av  $x$ ) instansierer en egenskap  $P$ . For det første forteller prinsippet oss at hvis  $x$  skal utøve disse kausale kreftene ved  $t$ , så må vi ha at  $P(x)$  ved  $t$ . For det andre forteller prinsippet oss at det faktum at  $P(x)$  forårsakes ved  $t$ , ikke er nok til at  $P(x)$  ved  $t$ . Moralen er altså at  $x$  ikke kan utøve kausale krefter som er avhengige av  $P(x)$  ved  $t$ , med mindre det at  $P(x)$  var sant allerede *før*  $t$ .

I tilfellet synkron refleksiv DC skal  $M(x)$  ved  $t$ , forårsake  $P_j(a_j)$  ved  $t$ , der  $a_j$  er en mikrokonstituent av  $x$ . To logiske slutninger avdekker at dette er umulig: (1) Anta prinsippet og at  $M(x)$  forårsaker  $P_j(a_j)$  ved  $t$ . Da har vi ikke ennå at  $P_j(a_j)$  ved  $t$ . (Jf. det andre poenget i prinsippet.) (2) Pga. prinsippet og at det ikke er tilfellet at  $P_j(a_j)$  ved  $t$ , så kan det heller ikke være tilfellet at  $M(x)$  forårsaker  $P_j(a_j)$  ved  $t$ , slik vi antok. (Jf. det første poenget i prinsippet.) Kim sier at  $x$  ikke kan ha  $M$  ved  $t$ , fordi emergensbasen til  $M$  er forsvunnet.<sup>237</sup> Essensen i argumentet er altså at  $M$  er avhengig av  $P_j$  for å forårsake  $P_j$  ved  $t$ , og siden vi ikke har  $P_j$  ved  $t$  (fordi  $M$  forårsaker  $P_j$  ved  $t$ ), så kan heller ikke  $M$  forårsake  $P_j$  ved  $t$ . Metaforisk kan vi si at synkron refleksiv DC ”slår bena under seg selv,” ved å forsøke å forårsake sin egen betingelse. Dette bekrefter at argumentet utnytter den asymmetriske avhengighetsrelasjonen emergentister er forpliktet til. Kim ønsker å unngå enhver form for resiprok avhengighet mellom det mentale og det fysiske. Det er viktig for hans fysikalistiske prosjekt at fysikk har forrang fremfor psykologi.

På den andre siden omhandler prinsippet bare tilfeller der DC skjer ved samme tidspunkt som ”upward determination,” så det berører ikke *diakron* refleksiv DC:

---

<sup>234</sup> Ibid. s.29.

<sup>235</sup> Ibid. s.28.

<sup>236</sup> Ibid. s.29

<sup>237</sup> Ibid. s.29

*W*'s having *M* at *t* causes *a<sub>j</sub>* to have *Q* at *t* +  $\Delta t$ . But *a<sub>j</sub>*'s having *Q* at *t* +  $\Delta t$  is not part of the basal conditions out of which *M* emerges in *W* at *t*; so there can be no problem of circular reciprocal causation/determination.<sup>238</sup>

Kim understreker at dette er uproblematisk og forekommer ofte, og hevder at Sperrys eksempler på DC – vi skal se på et av dem nedenfor (8.2-8.3) – ved nærmere undersøkelse viser seg å være av typen diakron DC. Men han hevder også at eksemplene ikke viser at ”emergent causation” er noe ”special and unique.”<sup>239</sup>

Så langt mener Kim altså å ha vist at synkron refleksiv DC er inkoherent, mens diakron refleksiv DC og nonrefleksiv DC er uproblematisk, men formodentlig ikke ”spesielt” nok til å tilfredsstille emergentistens behov. Hans eksempel med vasen innebærer da også DC fra en resultant, ikke emergent egenskap. Kim vil derfor se på DC fra mentale egenskaper, under forutsetning av at disse er emergente.<sup>240</sup> Han utelukker *synkron* refleksiv DC fra mentale egenskaper med henvisning til argumentet vi allerede har sett, og hevder at hverken Sperry eller klassiske emergentister egentlig trenger denne typen DC.<sup>241</sup> Derne utelukker han *diakron* refleksiv DC fra mentale egenskaper ut fra en variant av eksklusjonsproblemet.<sup>242</sup> Denne varianten er strukturelt og premissivt likt superveniensargumentet, bortsett fra at det er formulert ut fra emergens og ikke superveniens. Superveniensargumentet skal, som vi så, vise at egenskaper som ikke lar seg redusere, er kausalt impotente. Kims konklusjoner blir derfor: (1) Emergente egenskaper (som *ex hypotesi* er irreducibile) er kausalt impotente og eksplanatorisk ”largely useless”. (2) Mer generelt er det slik at ”[...] higher-level properties can serve as causes in downward causal relations only if they are reducible to lower-level properties.”<sup>243</sup> (2) er selvsagt paradoksalt: Kim foreslår derfor å gi DC en *konseptuell* fortolkning der DC skjer innenfor et hierarki av begreper og beskrivelser, ikke et hierarki av fenomener og egenskaper i verden.<sup>244</sup>

## 8 En vei utenom Kims argumenter mot ”downward causation?”

### 8.1 Innledning

Vi kjenner nå Kims posisjon godt nok til å begynne å sy sammen et emergentistisk forsvar mot den. Det vi vet er at to krav stilles til vår modell for DC: (1) Vi må unngå

---

<sup>238</sup> Ibid. s.30.

<sup>239</sup> Ibid.

<sup>240</sup> Ibid. s.31.

<sup>241</sup> Ibid. s.30.

<sup>242</sup> Ibid. s.32-33.

<sup>243</sup> Ibid. s.33.

<sup>244</sup> Ibid.

supervenienensargumentets konklusjon. For å få til dette har jeg foreslått at vi må akseptere brudd på det fysiske domenes kausale lukkethet. Men vi må også unngå de deler av Kims argumentasjon som er uavhengig av prinsippet om kausal lukkethet. Jeg har foreslått å bygge videre på arbeidsdelingsmodellen for DC, der den mentale egenskapen og dens fysiske superveniensbase til sammen utgjør en tilstrekkelig årsak. (2) Vår modell for DC må enten ikke innebære synkron refleksiv DC, eller vi må finne en feil i Kims argument for umuligheten av dette.

Jeg vil begynne med krav (2), og diskutere synkron refleksiv DC. Dette skyldes ikke at emergentisme står og faller med denne typen DC, eller at jeg absolutt vil redde den. Det kan godt hende at Sperrys eksempler må fortolkes som *diakron* DC. Det skyldes derimot at Kim misforstår Sperrys modell for DC, og det å avdekke denne misforståelsen gir oss kimen til en mer presis arbeidsdelingsmodell. Dermed kan vi lære noe om hvordan vi skal tilfredsstillere krav (1).

Det vil bli klart at denne arbeidsdelingsmodellen kun er en *skisse* av en mulig måte å møte Kims argumenter på. Mye presisering må gjøres før den – eventuelt – kan kalles en klar og forsvarlig tese. Til tross for at dette kapittelet derfor argumentativt sett er noe løsere enn tidligere kapitler, mener jeg at jeg i det minste peker på noen ideer som kan være fruktbare områder for videre undersøkelse.

## 8.2 Sperrys hjul I: "Downward causation" uten interferens med mikrostrukturen

Jeg vil argumentere for at Kim misforstår (eller ser bort fra, eller overser) et viktig premiss for Sperry, nemlig at DC ikke interfererer med objektets mikrostruktur. Følgelig misforstår Kim Sperrys mekanisme for DC. Jeg vil basere min fremstilling av Sperry på artiklene "Macro-Versus Micro-Determinism" og "In Defense of Mentalism and Emergent Interaction."<sup>245</sup>

Kim har aldeles rett i at Sperrys teori inneholder de to premissene (i) DC og (ii) "upward determination": Sperry snakker om "[...] downward causal control of lower by higher forces, that is, control of the subatomic by the atomic, the atomic by the molecular, the molecular by the cellular, etc. [...]."<sup>246</sup> På den andre siden innrømmer han at komplekse heler er "[...] largely micro-determined by the laws and properties of the pre-existent parts (among other things) [...]."<sup>247</sup>

---

<sup>245</sup> Sperry (1986) og (1991). Timothy O'Connor ((1994) s.101) anbefaler disse artiklene som klare formuleringer av Sperrys teori. Dessuten går Sperry her eksplisitt inn for å oppklare misforståelser av sin teori: Sperry (1986) s.265 og (1991) s.223

<sup>246</sup> Sperry (1986) s.265.

<sup>247</sup> Ibid. s.267



Imidlertid ser Kim som sagt bort fra et *tredje* premiss som kvalifiserer premiss (i) om DC: (iii) *DC interfererer ikke med objektets mikrostruktur*.<sup>248</sup> Ved å se bort fra dette misforstår Kim også Sperrys premiss (i) om DC, og hans argument rammer ikke Sperry.

Hvordan skal så noninterferenspremisset forstås? Sperrys mest berømte eksempel på DC er et hjul som ruller nedover en bakke. Delene i hjulet følger andre kurver enn de ville gjort dersom de ikke hadde inngått i helheten hjulet utgjør:

[...] a wheel rolling downhill [...] is a direct, simple, objective, physical example of macro-causation illustrating the universal principle of how the emergent properties of an entity as a whole exert *downward causal control* over the parts and the trajectories of the parts through space and time *without interfering with the causal interactions of the subentities at their own lower levels*.<sup>249</sup>

Påstanden (i) om DC fra hjulets makroegenskaper er her eksplisitt kombinert med kvalifiseringen (iii) om at DC ikke interfererer med interaksjoner eller prosesser på hjulets mikronivå. Sperry er litt mer presis i (1991) artikkelen:

A molecule within the rolling wheel, for example, though retaining its usual inter-molecular relations within the wheel, is at the same time, from the standpoint of an outside observer, being carried through particular patterns in space and time determined by the over-all properties of the wheel as a whole. There need be no “reconfiguring” of molecules relative to each other *within the wheel itself*. However, *relative to the rest of the world* the result is a major “reconfiguring” of the space-time trajectories of all components in the wheel’s infrastructure.<sup>250</sup>

Kim snakket om et hele  $W$ , med deler  $a_1, \dots, a_n$ , organisert i relasjonen  $R$ . Jeg vil på grunnlag av sitatet ovenfor våge påstanden om at noninterferens betyr at: *DC fra den emergente egenskapen  $M(W)$  ikke forandrer (“rekonfigurerer”)  $R$ , atomære egenskapsinstansieringer  $P_j(a_j)$  som  $M$  er avhengig av, eller eventuelle kausale prosesser mellom  $a_j$ ’ene*. Både delene og relasjonene internt i objektet forblir uforstyrret av DC. Denne lesningen synes å bekreftes når Sperry sier at DC er ”indirect”: DC er ”a superimposed intervention in events, not within, nor relative to things within the system itself, but *relative to the rest of the universe*.”<sup>251</sup> Jeg antar at frasen ”not within” i denne setningen omhandler atomære egenskapsinstansieringer ved hjulets deler, mens ”relative to things within” omhandler interne relasjoner i hjulet, og dermed spesielt  $R$ .

---

<sup>248</sup> Jeg ble oppmerksom på de relevante artiklene og dette premisset gjennom en artikkel av O’Connor. O’Connor tillegger Sperry premisset (O’Connor (1994) s.102), men avdekker ikke misforståelsen som oppstår hvis man ser bort fra det, og gjør lite for å presisere det. Kim er formodentlig klar over at premisset finnes hos Sperry, fordi han er klar over O’Connors artikkel. (Kim (1999) s.32.n.35) Imidlertid er det angivelig vanlig å betrakte Sperrys doktrine som uklar. (Se f.eks. Stephan (1992) s.43.) Kim sier faktisk også at han ikke ønsker å utelukke andre mulige lesninger av Sperry. (Kim (1999) s.27.) Vi bør i alle fall tillegge de to artiklene jeg bruker stor vekt, fordi Sperry som nevnt her eksplisitt går inn for å oppklare misforståelser.

<sup>249</sup> Sperry (1986) s.266, mine kursiveringer.

<sup>250</sup> Sperry (1991) s.230. Sitatet finnes også hos O’Connor. (O’Connor (1994) s.101)

<sup>251</sup> Sperry (1991) s.225.

Kims argument forutsatte at det som forårsakes ved synkron refleksiv DC fra M, er egenskaper som M er avhengige av, og hevdet at denne formen for DC derfor er umulig. Noninterferenspremisset benekter at det er slik DC fungerer. Argumentet berører derfor ikke Sperrys eksempel.<sup>252</sup>

Ideen om noninterferens har en viss affinitet til McLaughlins poeng om at konfigurasjonelle krefter (dvs. fundamentale krefter som bare kan utøves av bestemte strukturer) er konsistent med fysikkens lover. (5.3) (Derfor mener McLaughlin selvsagt også at Sperrys DC er logisk konsistent, men dog empirisk sett problematisk fordi han mener det ikke ser ut til å finnes noen konfigurasjonelle krefter.)<sup>253</sup> Sperrys noninterferenspremiss skaper derfor et mulig rom for at flere teorier som benytter seg av vesensforskjellige lover og prinsipper, kan være *om* den samme verden: Han sier faktisk eksplisitt at nye lover oppstår når nye heler oppstår.<sup>254</sup>

Men det er én ting å si at det finnes egenskaper ved W som M ikke er avhengig av, og som M følgelig kan kausalpåvirke. En ganske annen ting er hvilke egenskapsinstansieringer M faktisk forårsaker, og her ser det ut til at Sperry skylder oss en mer positiv historie. Jeg tror imidlertid at denne historien er implisitt i sitatene vi har sett på: For la oss tenke oss to forskjellige koordinatsystemer der vi kan representere posisjonene til hjulets deler over tid, og dermed også relasjonene mellom delene. Det første koordinatsystemet representerer bare delenes posisjoner internt i, og relativt til, hjulet. I dette koordinatsystemet vil delenes posisjoner, innbyrdes relasjoner, og eventuelle prosesser mellom delene være konstant. DC *à la* Sperry medfører ingen forandringer vi kan representere i dette systemet. I det andre koordinatsystemet representerer vi imidlertid også delenes posisjon relativt til den utenforliggende verden. Her *vil* vi observere forandringer i relasjoner både internt og relativt til resten av verden når hjulet ruller.

To molekyler A og B kan f.eks. relativt til det første systemet forbli i samme relasjon til hverandre, mens de relativt til det andre systemet kan skifte fra "A er til høyre for B" til "B er ovenfor A," og likeledes relativt til punkter i verden. Hva mer er: Disse endringene ville ikke skjedd hvis ikke hjulet hadde visse makroegenskaper knyttet til dets form, og endringene

---

<sup>252</sup> Om hjulet – siden det er i bevegelse – virkelig er et eksempel på *synkron* refleksiv DC, er selvsagt et annet spørsmål. Hvis man ville forsvare dette, kunne man kanskje si følgende: Hjulets emergente egenskap bestemmer kontinuerlig – dvs. til enhver tid – delenes posisjon der og da.

<sup>253</sup> McLaughlin (1992) s.90-91.

<sup>254</sup> Sperry (1986) s.267. Sperry antyder selv at hans teori kan ha vidtrekkende konsekvenser for hva som betraktes som vitenskapelig: Emergente krefter kan inkludere mentale, politiske, religiøse og andre sosiale krefter, så: "Scientific descriptions are no longer in utter conflict with traditional humanist views and values or the search for meaning." (Sperry (1986) s.269) Når det gjelder dette aspektet ved Sperrys teori, vil formodentlig frasen "for godt til å være sant" ligge snublende nær for flere.

kan antas å skje med nomologisk nødvendighet. Med andre ord synes det ikke urimelig å snakke om en *slags* kausale sammenhenger, der det som forårsakes er en art relasjonelle egenskaper.

Hvis vi går tilbake til Sperrys tekst igjen, finner vi at han vektlegger nettopp at det relasjonelle, spatio-temporale aspektet ved et objekt har en bestemmende rolle, i tillegg til objektets mikroegenskaper:

To help explain why a strict micro-determinism cannot fully account for the wheel example or for other emergent properties, I have repeatedly stressed the important causal role of nonmaterial space-time, pattern, or form factors and suggested that it is helpful to view any entity as not just a collection of material parts but as a mass-energy-space-time manifold built of space-time components as well as of matter. The emergent properties of the entirety and the laws for its causal interactions are determined by the spacing and timing as well as by the properties of the parts themselves.<sup>255</sup>

Emergente egenskaper bestemmes altså også av komplekse relasjonelle spatio-temporale egenskaper, og ser ut til å forårsake nettopp relasjonelle egenskaper av denne typen.

Men merk at han *ikke* sier at de emergente egenskapene *er* slike relasjonelle egenskaper. Dette sitatet kan derimot tyde på at de emergente egenskapene er noe mer enn egenskapen å være en bestemt strukturell ordning av visse deler.<sup>256</sup> At dette skillet mellom struktur og emergente egenskaper lar seg opprettholde er selvsagt avgjørende for vårt formål, og vil også være det i forbindelse med Van Gulicks modell nedenfor: De emergente egenskapene må være noe mer enn objektets struktur, i den forstand at de ikke kan utledes fra viten om objektets struktur og dets delegenskaper. Dersom dette skulle vise seg å *ikke* holde, må vi antagelig finne oss i en svakere form for emergens, og dermed også en svakere form for autonomi for psykologien. Forhåpentligvis ville vi likevel kunne dra nytte av noen trekk ved Sperrys og Van Gulicks modeller for DC.

### 8.3 Sperrys hjul II: Arbeidsdeling mellom mentale og fysiske egenskaper

I superveniensargumentet utelukket Kim modellen for DC der den mentale egenskapen M og dens superveniensbase P til sammen utgjør en tilstrekkelig årsak. (MPW 44.) Dette gjorde han bl.a. pga. antagelsen om at M alene skulle være tilstrekkelig. Jeg vil stille spørsmålstegn ved denne antagelsen, og kort si noe om hvordan Sperrys DC modell kan belyse hvordan M og P kan samarbeide som årsaker.

---

<sup>255</sup> Sperry (1986) s.266.

<sup>256</sup> Selv om jeg ikke kategorisk vil utelukke lesninger av Sperry der de emergente egenskapene *er* slike strukturelle egenskaper. O'Connor mener som jeg at "non-structurality" er et viktig trekk ved emergente egenskaper, (O'Connor (1994) s.97), og formodentlig også ved Sperrys emergensbegrep slik han fortolker det.

Dersom en mental egenskap M kunne gjøre en kausal jobb uavhengig av dens fysiske superveniensbase P, så ville vi ha garantert en sterk autonomi for psykologien. Men innenfor et fysikalistisk perspektiv kan det synes urimelig at mentale egenskaper forårsaker andre egenskaper uavhengig av de nevrale egenskapene de er superveniente på. De nevrale egenskapene er selvsagt betingelser for mentale egenskaper, men det ville ikke være urimelig å forvente at de *også* skulle gjøre et direkte kausalt bidrag. Slike betraktninger kan motivere tanken på å gi opp M som tilstrekkelig årsak, og konstruere en arbeidsdelingsmodell der M og P til sammen gjør den kausale jobben.

Sperrys teori kan brukes til å se mulige mekanismer for slike modeller. For hjulets mikrostruktur betinger de emergente egenskapene, og gir samtidig et avgjørende bidrag til prosessen å rulle nedover. Så den subveniente egenskapen er med på å gjøre jobben. Dette er spesielt klart fordi Sperry ikke hevder at makrodeterminisme skal erstatte mikrodeterminisme, men supplere den: Han snakker om et paradigmeskifte i vitenskapshistorien på 1970-tallet fra mikrodeterminisme, til et fokus på makronivå der makrovitenskaper har genuin eksplanatorisk verdi.<sup>257</sup>

The recent overthrow in psychology of this seemingly impregnable position [that micro sciences can give a complete explanation of brainprocesses, without referring to consciousness] succeeded, not so much by negating or displacing micro-determinism, but rather by supplementing it. Micro-determinism is retained but is held to be incomplete, insufficient. The properties, forces and laws of micro-events are shown to be encompassed and superseded, not disrupted, by the properties, forces, and laws at macro-levels.<sup>258</sup>

Uavhengig av om et slikt paradigmeskifte virkelig har funnet sted, så synes det trygt å si at Kim befinner seg innenfor et mikrodeterministisk paradigme, og at dette styrer tankegangen i hans eksklusjonsbaserte argument. Hans fokus på det subveniente nivået kan således ha gjort ham blind for muligheter der det superveniente nivået spiller en rolle.

Sperrys teori trenger selvsagt å klargjøres og nyanseres for å møte åpenbare innvendinger. Jeg vil se på noen slike problemer nedenfor.

#### **8.4 Problemer med arbeidsdelingsmodellen. Van Gulicks mønstre**

Et åpenbart problem med Sperrys modell for DC er at hjuleksempelen er et eksempel på en resultant, ikke emergent egenskap. Kim sier at egenskapene som er involvert i Sperrys

---

<sup>257</sup> Han gir en rekke referanser for å dokumentere paradigmeskiftet i Sperry (1991). (Spes. s.237-242). Et mulig problem med påstanden om dette skiftet er at Sperry regner f.eks. "cognitive science" til det nye paradigmet. McLaughlin hevder (kanskje plausibelt) at "cognitive science" ikke forutsetter "new [emergent] forces", men komputasjonalisme. (McLaughlin (1992) s.91.n.48) Dette tyder på at McLaughlin eller Sperry har misforstått hva visse moderne vitenskapelige praksiser består i.

<sup>258</sup> Sperry (1986) s.268.

eksempler på DC er av samme type: Bevegelse forårsaker bevegelse.<sup>259</sup> Hvorvidt dette er en korrekt beskrivelse av hjuleksempelen, er tvilsomt: For min diskusjon ovenfor tyder på at det ikke er bevegelse, men en eller annen egenskap som er betinget av hjulets spatio-temporale form som – ifølge Sperry – forårsaker bevegelsen. (Husk Sperrys poeng om at ”spacing” og ”timing” er en bestemmende faktor ved siden av delenes egenskaper.) På den andre siden er det tilsynelatende ikke noe som hindrer en deduksjon av denne egenskapen fra viten om hjulets mikrostruktur og generelle bevegelseslover. Egenskapen må derfor være resultant ifølge vår definisjon.

O'Connor påpeker at emergens ifølge Sperry er svært utbredt også innen fysikk og kjemi, hvor vi har gode grunner til å tro at egenskaper er resultante. Han forklarer dette ved at Sperry har et ”highly novel” emergensbegrep.<sup>260</sup> Dette er antagelig riktig, men hjelper oss ikke stort, gitt vårt mer klassiske emergensbegrep. Men dette behøver heller ikke være et avgjørende problem:

For det første påpeker O'Connor at vi ifølge Sperry ikke kan kjenne relasjonene mellom et objekts komponenter bortsett fra i det den kaller ”ultra simple systems”. Systemer kan også ha komplekse ”dynamic infrastructures”. Dagens matematikk og vitenskap er ikke tilstrekkelige for å fange inn disse egenskapene.<sup>261</sup> Hjulet er formodentlig et slikt ultra enkelt system, og dynamikken mellom delene i hjulet må være beskjeden i forhold til dynamikken i nevrofysiologiske systemer. Derfor er det altså tilsynelatende også for Sperry en epistemisk kløft som gjør emergente egenskaper utilgjengelige fra viten om mikronivået. For det andre – og kanskje viktigst – kan vi overføre Sperrys noninterferenspremiss til egenskaper vi mener er emergente, selv om vi skulle mene at noen av Sperrys eksempler er eksempler på resultante egenskaper. Vi kan likevel trekke nytte av Sperrys modell for DC.

Et annet problem er hvorvidt DC uten interferens er sterkt nok til å telle som kausalitet i det hele tatt. Kan vi virkelig si at egenskaper som på uklart vis er betinget av bl.a. ”spacing” og ”timing” av hjulets deler, *forårsaker* at delene følger bestemte kurver? Vi må merke oss at vi kanskje ikke bør forvente kausalrelasjoner i den forstand vi er vant til å møte dem i

---

<sup>259</sup> Kim (1999) s.31.

<sup>260</sup> O'Connor (1994) s.101 med referanse til bl.a. Sperry (1986) s.266 og (1991) s.225 & 229.

<sup>261</sup> Sperry (1991) s.235, O'Connor (1994) s.101. Dette *kan* tyde på at Sperry har et teoreti-relativt emergensbegrep. Men fordi Sperry her argumenterer mot at man *prinsipielt* kan kjenne de komplekse egenskapene som betinger den emergente egenskapen, så tror jeg han må mene at relasjonene mellom delene er utilgjengelig for mennesket i en absolutt forstand. Men en laplaceansk demon ville kanskje kunne kjenne relasjonene, og dermed dedusere den emergente egenskapen.

filosofien: Kim tror at Sperry er mer interessert i avhengighetsrelasjoner i vid forstand enn i spesifikke filosofiske begreper om kausalitet.<sup>262</sup>

Stephan foreslår i stedet en svakere reformulering av Sperrys teori:

The same structures which determine the emergent macrophenomena *restrict the possibilities of the parts* they are constituted of. This view fits the examples given by Sperry, as there are the wheel rolling downhill or the eddy forcing waterdrops to swirl around. If one accepts this view there is no problem concerning the mechanisms [for DC]. However, Sperry would have to give up his idea of psycho-physical interactionism, [...].<sup>263</sup>

Stephan mener altså at Sperrys eksempler ikke støtter opp om noe så sterkt som (kausal) interaksjonisme. Men i lys av noninterferensprinsippet kan vi se at Stephan-sitatet antagelig ikke er en *svakere* variant av Sperrys teori, men en nokså korrekt beskrivelse av den. (Stephan forstår tilsynelatende Sperrys DC på samme måte som Kim, og er blind for noninterferens-premisset.)<sup>264</sup> For det er klart at hjulets makroegenskaper som Stephan sier begrenser delenes muligheter, i den forstand at de ikke kan følge de kurver de *ville fulgt* hvis de *ikke* hadde inngått i hjulet. Men beskrivelsen ”begrenser mulighetene” er ikke helt tilstrekkelig, siden vi må si at delene også har muligheter i hjulet som de *ikke* ville hatt alene. Så DC *a lá* Sperry utøver både en positiv og negativ påvirkning på delenes kausale potensialer: DC gir og tar muligheter.<sup>265</sup>

---

<sup>262</sup> Kim (2000) s.314.

<sup>263</sup> Stephan (1992) s. 45, min kursivering.

<sup>264</sup> Stephan (1992) s.44

<sup>265</sup> Det har nylig skjedd en del på dette området, og mer subtile og interessante varianter av DC, som ligner Stephans forslag om at hjulet *begrenser* delenes muligheter, er blitt formulert: Biologene El-Hani og Emmeche diskuterer muligheten av å fortolke DC nettopp som at et komplekst systems form legger begrensninger på delenes muligheter. De fortolker Sperry i tråd med denne stephanske modellen, uten, som meg, å legge til at delene også får *nye* muligheter. (El-Hani & Emmeche (2000) s.262-266. Se også Emmeche, Köppe & Stjernfeldt (2000) s.24.) Det er selvsagt også riktig å si at det som skjer alt i alt er en begrensning av de mulighetene som de isolerte delene har *logisk* sett. I denne oppgaven foretrekker jeg likevel formuleringen ”DC gir og tar muligheter,” fordi den får frem at nettopp det at delene er *i* systemet gjør en positiv forskjell i forhold til hva delene *faktisk* kunne gjort alene.

Den mekanismen for DC jeg skisserer her avviker også mer substansielt fra El-Hani og Emmeches modell: (1) Deres syn er mer *reduktivt*, fordi emergente egenskaper i følge dem kan være *strukturelle* egenskaper, og ikke bare være betinget av slike. Kanskje kan de også være forutsigbare. (El-Hani & Emmeche (2000) s.266.) Dette kan selvsagt også tenkes at de relevante egenskapene i en stephansk modell for DC faktisk vil vise seg å være strukturelle og resultante/forutsigbare. Men en slik løsning ville være av en mer reduktiv natur, og psykologien ville ikke få den autonomien vi ønsket. El-Hani og Emmeche innrømmer faktisk at deres løsning i en viss forstand er reduktiv. (Ibid. s.268-269.) (2) El-Hani & Emmeche beskriver sin modell for DC som aristotelisk (eller kvasi-aristotelisk) ”formal causation,” ikke som alminnelig ”efficient causation.” (Ibid. s.267.) Fordi DC altså er en spesiell form for kausalitet, mener de (ibid.) at det fysiske kausale lukkethet *ikke* brytes. Arbeidsdelingsmodellen jeg skisserer hevder imidlertid at den emergente årsaken inngår som en nødvendig ikke-fysisk del av en tilstrekkelig ”efficient cause,” og at lukketheten derfor brytes. (Jf. MPW 44.) Jeg vil ikke påstå at det er åpenbart om det å ”begrense delenes muligheter” skal forstås som arbeidsdeling eller som ”formal causation.” Likevel tror jeg det er mulig å presse El-Hani & Emmeches modell: Enten er den emergente egenskapen en nødvendig betingelse for at de fysiske kausale kjedene tar den veien de gjør, eller så er den det ikke. Hvis den er nødvendig må en komplett fysisk historie henvise til den for å forklare kausalprosessen, og dette synes som et brudd på lukkethetsprinsippet ideologi. I den grad egenskapen derimot *ikke* er nødvendig, trues dens eksplanatoriske relevans og realitet som egenskap, i det minste under de realistiske forutsetningene i

Vi har nå kommet til et punkt hvor vi kan se en interessant affinitet mellom Sperry og Van Gulick. Jeg kan imidlertid her bare kort oppsummere Van Gulicks interessante ideer omkring DC: Spesialvitenskaper klassifiserer objekter og hendelser etter kausale roller ved å plukke ut stabile, gjentakende mønstre i verden. Hendelsene og objektene er selvsagt sammensatt av fysiske konstituentene, men spesialvitenskapenes predikater plukker ut nettopp *mønsteret konstituentene er organisert i*.<sup>266</sup>

Uten å nevne Sperry har Van Gulick et argument for at slike mønstre (og ikke bare materien som utgjør dem) er reelle og har kausale krefter: For det første er de gjentakende og stabile ("recurrent and stable") trekk ved verden. For det andre kan de være stabile på tross av utskiftning av eller forandringer i de underliggende delene. (Han nevner en tornado som eksempel.) For det tredje kan de (som DNA-mønstre) være selv-oppretholdende og selv-reproduserende. For det fjerde og (mest interessant) kan mønstrene "affect which causal powers of their constituents are activated or likely to be activated. [...] For example, the activity of a reagent can be affected by the presence of a catalysing enzyme that forms a composite with the reagent."<sup>267</sup> For det femte kan denne fjerde mekanismen bidra til selv-oppretholdelsen av mønsteret.<sup>268</sup>

I lys av Sperry-lesningen vi kom frem til ovenfor, bringer følgende sitat frem en slående likhet mellom Sperrys og Van Gulicks modeller for DC:

Taken together these five points illustrate that higher-order patterns can have a degree of independence from their underlying physical realizations and can exert what might be called downward causal influences without requiring any objectionable form of emergentism by which higher-order properties would alter the underlying laws of physics. Higher-order properties act by the *selective activation* of physical powers not by their *alteration*.<sup>269</sup>

Van Gulicks "selektiv aktiverings" modell av delenes kausale potensialer gir oss en generell beskrivelse av hvordan komplekse dynamiske systemer kan oppføre seg. I tillegg til dette har jeg i forlengelsen av Sperrys hjuleksempel foreslått at delene også får nye potensialer, potensialer de ikke ville hatt utenfor systemet. Gitt dette kausale samarbeidet

---

denne oppgaven. (En annen sak er at El-Hani og Emmeches egenskaper kanskje vil kvalifisere som fysiske i Kims vide forstand av ordet. I så fall vil *Kim* antagelig ikke kritisere dem ut fra lukkethet i det hele tatt.) Uansett synes det trygt å si at debatten om alternative modeller for DC er langt fra å være avsluttet.

<sup>266</sup> Van Gulick (1993) s.250

<sup>267</sup> Ibid. s.251

<sup>268</sup> Ibid. s.252.

<sup>269</sup> Ibid. s.252. Van Gulick erklærer seg vel å merke ikke som emergentist her, han er en selverklært "teleo-funksjonalist" (Van Gulick (1992) s.161) hvilket synes å bety at han er en slags funksjonalistisk pragmatist. Men mange av hans ideer har klare likhetstrekk med emergentisme. Jeg forutsetter imidlertid ikke at han nødvendigvis ville si seg enig i emergentismens epistemologiske påstander om uforutsigbarhet o.l., selv om vi har sett at han ikke tror det kan finnes ryddige og lovmessige sammenhenger mellom psykologi og fysikk. (1.3) Det er hans DC mekanisme som er interessant her og nå.

mellom deler og den emergente egenskapen er det klart at begge er nødvendige ingredienser i årsaken. I spesialtilfellet av mentale egenskaper må derfor modellene sies å være av den typen der den mentale egenskapen og dens fysiske superveniensbase til sammen er en tilstrekkelig årsak. Dvs. vi har en arbeidsdelingsmodell.<sup>270</sup> I en viss forstand foregår altså DC fra en mental egenskap via, og sammen med, dens superveniensbase.

Alt i alt synes det derfor som om Sperry og Van Gulick har skissert konsistente mekanismer for DC som har en viss plausibilitet. Mekanismene kan sies å kombinere den emergentistiske ideen om nye, konfigurasjonelle krefter med noninterferenspremisset og med at mikro- og makronivå samarbeider som årsaker. Mekanismene berøres dermed ikke av Kims argumenter.

### 8.5 Platonisk reifisering? Intuisjoner og vitenskapelig praksis

Men det er en ting å skissere en plausibel generell mekanisme, en ganske annen ting å ha en holdbar teori. For det er klart at arbeidsdelingsmodellen vil stå i motsetning til manges intuisjoner om at *det virkelige eller reelle er de fysiske konstituentene* i Van Gulicks mønstre, og at å betrakte mønsteret som noe ”over and above its strictly physical instantiations would be an exercise in Platonic reification.”<sup>271</sup>

Mekanismene som er skissert er generelle og vage, og *må* være det fordi de skal passe vesensforskjellige systemer. Jeg tror likevel det kan være nyttig å prøve den antiplatoniske intuisjonen mot et konkret eksempel: Glaseologer studerer isbreer, og mer generelt kan man studere andre frosne vannansamlinger som fosser. Disse objektene er stabile gjentakende mønstre av organiserte vannmolekyler og urenheter. De er i noen grad selvoppretholdende (kan magasinere kulde) og overlever utskiftning av deler (delvis smelting og ny frysing). Prosesser som trykk og strekk spiller en rolle, og kan påvirke mikrostrukturen.

Van Gulick og Sperry vil tilsynelatende be oss se på våre frosne objekter som et mønster i tillegg til en ansamling av vannmolekyler. Imidlertid tror jeg at manges intuisjoner når de tenker på dette eksempelet vil gå i retning av at molekylene er kausalt reelle i en forstand mønsteret ikke er. Hva skyldes disse intuisjonene? En ting som synes å ligge til grunn, er den klassiske greske ontologiske tesen om at det som er avhengig av noe annet for

---

<sup>270</sup> Van Gulick beskriver imidlertid selv sin modell som en type kausal overbestemmelse. (Van Gulick (1993) s.255) Dette kan delvis skyldes at det eksklusjonsargumentet han diskuterer avhenger av et premiss som utelukker overbestemmelse. (Ibid. s.242.) Uansett synes det klart at han betrakter både mønsteret og delene som nødvendige for objektets kausale krefter (ibid. 250), så virkningen kan ikke betraktes som ”overbestemt” i betydningen ”har to tilstrekkelige årsaker.”

<sup>271</sup> Van Gulick (1993) s.251.



sin væren, er mindre virkelig. (Ideer om selv-tilstrekkelighet dukker ofte opp i fysikalistiske posisjoner, et eksempel er prinsippet om det fysiskes kausale lukkethet.) En annen mulig årsak er at vi kan være preget av et mekanistisk, og mer eller mindre mikrodeterministisk paradigme for vitenskapelig praksis, en forestilling som i sin mest primitive form ser på verden som et urverk. Uten å si noe om disse forestillingenes gyldighet synes det trygt å si at de i hvert fall delvis farver bildet av den fysiske verden, slik den fremstilles i skoleverket og i populærvitenskapelige sammenhenger.

Van Gulick er inne på noe lignende når han drøfter følgende innvending: "And physical properties are not mere stable recurrent patterns; they are the basic stuff of which all the patterns are made."<sup>272</sup> Van Gulick sier imidlertid at han vil skape *likhet* mellom fysikk og spesialvitenskaper ved å berøve fysikken for dens påstått spesielle status: "The physicalist emperor has no clothes."<sup>273</sup> Van Gulicks radikale løsning består i å foreslå at heller ikke fysiske egenskaper er noe mer enn stabile gjentakende mønstre:

Though it is a question whose answer requires more knowledge of physics than I possess, I believe it could be shown that all physical properties that enter into strict exceptionless laws are themselves nothing more than stable self-sustaining recurrent states of the quantum flux of an irreducibly probabilistic and statistical reality. If cosmologists are to be believed (and I believe them), our universe settled into these patterns very early on, some time within the first three minutes after the Big Bang. But their antiquity does not alter their status.<sup>274</sup>

Han sier at disse mønstrene tilnærmer deterministiske regulariteter, og sporer forestillingen om at de er mer virkelige, tilbake til dette:

I think we are often bewitched by their deterministic rigour, which produces in us an illusion of dealing with more than just a representation of some aspect of the organization of space-time. We feel that we are dealing with something more tangible, something more real and objective. Their determinism seduces us into seeing physical properties as determining the sequence of events in a way that no other properties do.<sup>275</sup>

Sperry på sin side hevder, som vi har sett, at hans kritikere har misforstått paradigmeskiftet fra et mikroreduksjonistisk ideal til hans eget ideal om at komplisert "spacing" og "timing" er en bestemmende faktor ved siden av delegenskaper. Et objekt bør derfor betraktes som et "mass-energy-space-time manifold."<sup>276</sup> Både Sperry og Van Gulick kan derfor på hver sin måte sies å representere et vitenskapsideal som vektlegger spatio-temporale mønstre som bestemmende faktorer.

---

<sup>272</sup> Van Gulick (1993) s.254.

<sup>273</sup> Van Gulick (1993) s.250. Dette betyr ikke at Van Gulick gir opp fysikalisme/materialisme. (Van Gulick (1992) s.161.) Han ser helt klart ut til å ville akseptere minimal fysikalisme. (Ibid. s.164.)

<sup>274</sup> Van Gulick (1993) s.254.

<sup>275</sup> Ibid.

<sup>276</sup> Sperry (1986) s.266.

Dersom de har rett ser det ut til at den antiplatoniske intuisjonen bunner i en antikvert oppfatning av vitenskapelig praksis der det eneste virkelige grovt sagt er små demokritske klumper av fysisk materie. *Om* de har rett, synes å avhenge bl.a. av hva som presist ligger i våre intuisjoner, hvordan fysikkens resultater skal fortolkes, og hvilke metafysiske forutsetninger om kausalitet som faktisk gjøres i fysisk og spesialvitenskapelig praksis. Kanskje har de rett. Kanskje ikke.

## 9 Epilog

I denne oppgaven har jeg forutsatt et bestemt filosofisk perspektiv: En betingelse har vært en sterk form for realisme, der det kreves at psykologiske forklaringer må vise til kausalt virksomme mentale egenskaper i verden. En annen betingelse har vært at denne verden er en grunnleggende fysisk verden: Minimal fysikalisme tilsier nemlig at alle objekter har fysiske egenskaper, og at mentale egenskaper er superveniente på fysiske egenskaper. På ønskelisten har det stått en *autonom psykologi*, forstått som en begrepslig praksis som er vesensforskjellig fra fysikk. Vi så imidlertid at den tette metafysiske relasjonen mellom mentale og fysiske egenskaper truet med å bringe psykologien innenfor fysikkens forklaringsmessige rekkevidde, og dermed sette autonomien i fare.

Jeg har altså i oppgaven gått med på svært mange av Kims metafysiske og realistiske forutsetninger, og undersøkt hvilken metafysisk pris man må betale for *likevel* å kunne ha en autonom psykologi. Ved å gjøre en rekke metafysiske krumspring mener jeg å ha rimeliggjort at emergentisme kan skape rom for en slik psykologi, og videre at emergentisme er en konsistent (men dog ikke empirisk uproblematisk) doktrine.

De metafysiske krumspringene inkluderer bl.a. å gi opp ideen om at mentale årsaker er tilstrekkelige årsaker, emergente egenskaper hvis empiriske status er usikker og en mekanisme for såkalt "downward causation" som kanskje bare kan kalles "kausal" i en svært utvannet betydning av ordet. Det fysiskes kausale lukkethet er en annen ting vi har gitt opp. Prisen for psykologiens autonomi har helt klart blitt høy. Noen vil sikkert hevde at den er *for* høy. Kanskje burde man forventet seg dette gitt de sterke fysikalistiske og realistiske betingelsene jeg startet fra. Dette kan motivere noen til å revurdere disse betingelsene. Andre vil kanskje heller sette mindre på ønskelisten, og undersøke om man f.eks. kunne klart seg

med et svakere begrep om en kausalt relevant (eller kausalt virksom) egenskap slik bl.a. Fodor forsøker.<sup>277</sup>

Men det kan være et generelt antireduktivt poeng her, som mine metafysiske krumspring kun er et eksempel på: For Kims reduktive argument forutsatte at irreducible egenskaper skulle interferere med, eller forstyrre, det selv-tilstrekkelige fysiske domenet. Han argumenterte så for at dette var umulig: Eksplanatorisk interferens – at mentale forklaringer skulle forstyrre fysiske forklaringskomplekthet – var umulig pga. prinsippet om eksplanatorisk eksklusjon. Kausal interferens – at mentale egenskaper skulle forstyrre fysiske kausale kjeder – var også umulig pga. eksklusjonsprinsippet, prinsippet om det fysiske kausale lukkethet, og dessuten fordi de mentale egenskapene ikke kunne endre sin betingelse. De metafysiske krumspringene jeg har gjort, har hatt som sitt primære mål å vise at antireduksjonisme til en viss grad *ikke* forstyrrer det fysiske på den måten Kim forutsetter. Og dette kan være en mer generell antireduktiv forutsetning; for også Burge hevder som vi så at det mentale ikke forstyrrer det fysiske, men han hevder dette ut fra en helt annen metafysikk enn Sperry og Van Gulick.<sup>278</sup> Så en videre undersøkelse av antireduksjonisme kunne ta utgangspunkt i denne generelle ideen om noninterferens, og Kims forutsetning om at antireduksjonister er forpliktet til interferens.

Uansett: Den noninterferensbaserte løsningen jeg har skissert har vist seg å avhenge av intuisjoner og empiriske resultater, samt kausalrelasjonens og vitenskapelige praksisers natur. Disse må gjøres klarere før det kan vurderes ordentlig om den skisserte formen for emergentisme er en farbar vei. I tråd med dette er det kanskje passende å avslutte med en kommentar Fodor har gjort i en annen sammenheng. (Han regner seg forøvrig som en "impure philosopher"):

Pure philosophy thus reveals two ways in which [a question concerning mental content] might have an answer; and, having done so, washes its hands and departs. So like pure philosophy, to leave before the party starts.<sup>279</sup>

---

<sup>277</sup> Fodor (1989) s.143 og 146.

<sup>278</sup> Burge (1993) s.116, se (7.3).

<sup>279</sup> Fodor (1994) s.27.

## 10 Bibliografi

Alexander, Samuel (1920): *Space, Time and Deity*, vol.2. London: Macmillan.

Andersen, P.B, Emmeche, C., Finnemann, N.O. & Christiansen, P.V. (eds.) (2000):  
*Downward Causation, Minds, Bodies and Matter*, Aarhus University Press.

Beckerman, A., Flohr, H., Kim, J. (eds.) (1992): *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism*, Berlin: Walter de Gruyter.

Beckermann, Ansgar (1992): "Supervenience, Emergence, and Reduction" i Beckermann et. al. (1992).

Broad, C.D. (1925): *The Mind and Its Place in Nature*, London: Kegan Paul, Trench, Trubner & Co. Ltd.

Burge, Tyler (1993): "Mind-Body Causation and Explanatory Practice" i Heil & Mele (1993).

Caston, Victor (1997): "Epiphenomenalisms, Ancient and Modern", *The Philosophical Review*, Vol.106, No.3, July.

Churchland, Paul M. (1985): "Reduction, Qualia, and the Direct Introspection of Brain States", *The Journal of Philosophy*, vol. 82, issue 1.

Churchland, Paul M. (1988): *Matter and Consciousness. A Contemporary Introduction to the Philosophy of Mind*, Cambridge/Massachusetts: MIT Press.

Churchland, Paul M. (1995): *The Engine of Reason, the Seat of the Soul, A Philosophical Journey into the Brain*, Cambridge/Massachusetts: MIT Press.

Churchland, Paul M. (1996): "Folk Psychology" i Guttenplan, S. (ed.): *Companion to the Mind*, Oxford: Blackwells, gjengitt i Churchland & Churchland (1998). Mine referanser er til pagineringen i Churchland & Churchland (1998).

Churchland, Paul M. & Patricia S. (1990): "Intertheoretic Reduction: A Neuroscientist's Field Guide", *Seminars in the Neurosciences* 1990, 249-56, gjengitt i Churchland & Churchland (1998). Mine referanser er til pagineringen i Churchland & Churchland (1998).

Churchland, Paul M. & Patricia S. (1998): *On the Contrary, Critical Essays, 1987-1997*, Cambridge, Massachusetts, London, England: MIT Press.

Davidson, Donald (1963): "Actions, Reasons and Causes", *Journal of Philosophy* 60, gjengitt i Davidson (1980). Mine referanser er til pagineringen i Davidson (1980).

Davidson, Donald (1967): "Causal Relations", *Journal of Philosophy* 64, gjengitt i Davidson (1980). Mine referanser er til pagineringen i Davidson (1980).

Davidson, Donald (1969): "The Individuation of Events" i Rescher, N. & Reidel D. (eds.): *Essays in Honor of Carl G. Hempel*, Dordrecht-Holland: D.Reidel Publishing Company, gjengitt i Davidson (1980). Mine referanser er til pagineringen i Davidson (1980).

Davidson, Donald (1970): "Mental Events", i Foster, L. & Swanson, J.W (eds.): *Experience and Theory*, The University of Massachusetts Press and Duckworth, gjengitt i Davidson (1980). Mine referanser er til pagineringen i Davidson (1980).

Davidson (1974): "On the Very Idea of a Conceptual Scheme", *Proceedings and Addresses of the American Philosophical Association* 47, gjengitt i Davidson (1984).

Davidson, Donald (1980): *Essays on Actions and Events*, Oxford: Clarendon Press.

Davidson, Donald (1984): *Inquiries into Truth and Interpretation*, Oxford: Clarendon Press.

Davidson, Donald (1993): "Thinking Causes" i Heil & Mele (1993).

Davidson, Donald (1995): "Laws and cause" *Dialectica*, vol.49, no. 2-4.

Descartes, René (1992): *Meditasjoner over filosofiens grunnlag*, overs. Asbjørn Aarnes, Oslo: H. Aschehoug & Co. / Thorleif Dahls Kulturbibliotek.

El-Hani, Charbel Ninõ & Emmeche, Claus (2000): "On Some Theoretical Grounds for an Organism-centered Biology: Property Emergence, Supervenience and Downward Causation", *Theory in Biosciences*, vol. 119.

Emmeche, Claus, Køppe, Simo & Stjernfeldt Frederik (2000): "Levels, Emergence, and Three Versions of Downward Causation" i Andersen et.al. (2000).

Flohr, Hans (1992): "Qualia and Brain Processes" i Beckermann et. al. (1992).

Fodor, Jerry A. (1974): "Special Sciences, or The Disunity of Science as a Working Hypothesis", *Synthese* 28, 97-115, gjengitt i Fodor (1981). Mine referanser er til pagineringen i Fodor (1981).

Fodor, Jerry A. (1981): *Representations, Philosophical Essays on the Foundations of Cognitive Science*, Brighton: The Harvester Press.

Fodor, Jerry A. (1987): *Psychosemantics, The Problem of Meaning in the Philosophy of Mind*, Cambridge, Massachusetts: MIT Press.

Fodor, Jerry A. (1989): "Making Mind Matter More", *Philosophical Topics* 17, 59-79 gjengitt i Fodor (1990). Mine referanser er til pagineringen i Fodor (1990).

Fodor, Jerry A. (1990): *A Theory of Content and Other Essays*, Cambridge, Massachusetts: MIT Press.

Fodor, Jerry A. (1994): *The Elm and the Expert, Mentalese and Its Semantics*, Cambridge, Massachusetts: MIT Press.

Hansen, Carsten Martin (2000): "Between a Rock and a Hard Place: Mental Causation and the Mind-Body Problem", *Inquiry* vol.43, no.4, dec. 2000.

Heil, John & Mele, Alfred (eds.) (1993): *Mental Causation*, Oxford: Clarendon Press.

Hempel, C.G. & Oppenheim, P. (1948): "Studies in the Logic of Explanation" *Philosophy of Science* 15, 135-175.

Hume, David (1927): *An Enquiry Concerning Human Understanding And Selections from A Treatise of Human Nature*, Chicago: The Open Court Publishing Company.

Humphreys, Paul (1997): "Emergence, Not Supervenience", *Philosophy of Science* 64.

Jackson, Frank (1986): "What Mary Didn't Know", *The Journal of Philosophy*, vol.83, no. 2.

Kim, Jaegwon (1976): "Events as Property Exemplifications", i Brand, M. & Walton, D. (eds.): *Action Theory*, Dordrecht Holland: D. Reidel Publishing Co., gjengitt i Kim (1993c).  
Mine referanser er til pagineringen i Kim (1993c).

Kim, Jaegwon (1984a): "Concepts of Supervenience", *Philosophy and Phenomenological Research* 45, 153-176, gjengitt i Kim (1993c). Mine referanser er til pagineringen i Kim (1993c).

Kim, Jaegwon (1984b): "Epiphenomenal and Supervenient Causation", *Midwest Studies in Philosophy* 9, 257-270, gjengitt i Kim (1993c). Mine referanser er til pagineringen i Kim (1993c).

Kim, Jaegwon (1989a): "Mechanism, Purpose, and Explanatory Exclusion", *Philosophical Perspectives* 3, *Philosophy of Mind and Action Theory*: 77-108, ed. James E. Tomberlin, gjengitt i Kim (1993c). Mine referanser er til pagineringen i Kim (1993c).

Kim, Jaegwon (1989b): "The Myth of Nonreductive Materialism", *Proceedings and Addresses of the American Philosophical Association* 63, 31-47, gjengitt i Kim (1993c). Mine referanser er til pagineringen i Kim (1993c).

Kim, Jaegwon (1990): "Supervenience as a Philosophical Concept", *Metaphilosophy* 21, 1-27 gjengitt i Kim (1993c). Mine referanser er til pagineringen i Kim (1993c).

Kim, Jaegwon (1992): "Multiple Realization and the Metaphysics of Reduction", *Philosophy and Phenomenological Research* 52, 1-26, gjengitt i Kim (1993c). Mine referanser er til pagineringen i Kim (1993c).

Kim, Jaegwon (1993a): "Can Supervenience and 'Non-Strict Laws' Save Anomalous Monism?" i Heil & Mele (1993).

Kim, Jaegwon (1993b): "Postscripts on supervenience" i Kim (1993c).

Kim, Jaegwon (1993c): *Supervenience and mind, selected philosophical essays*, Cambridge University Press.

Kim Jaegwon (1993d): "The Non-Reductivist's Troubles with Mental Causation" i Heil & Mele (1993).

Kim, Jaegwon (1998a): *Mind in a Physical World, An Essay on the Mind-Body Problem and Mental Causation*, Cambridge, Massachusetts: MIT Press.

Kim, Jaegwon (1998b): *Philosophy of Mind*, Westview Press.

Kim, Jaegwon (1999): "Making Sense of Emergence", *Philosophical Studies* 95, 3-36.

Kim, Jaegwon (2000): "Making Sense of Downward Causation" i Andersen et.al. (2000).

McLaughlin, Brian P. (1992): "The Rise and Fall of British Emergentism" i Beckermann et. al. (1992).

McLaughlin, Brian P. (1993): "On Davidson's Response to the Charge of Epiphenomenalism" i Heil & Mele (1993).

Nagel, Ernest (1961): *The Structure of Science*, New York: Harcourt, Brace & World.

O'Connor, Timothy (1994): "Emergent Properties", *American Philosophical Quarterly*, vol. 31, number 2.

Oppenheim, Paul & Putnam, Hilary (1958): "Unity of Science as a Working Hypothesis", særtrykk fra Feigl, H., Scriven, M. & Maxwell G. (eds.): *Volume II., Minnesota Studies in the Philosophy of Science: Concepts, Theories and the Mind-Body Problem*, Minneapolis: University of Minnesota Press.

Pepper, Stephan (1926): "Emergence", *Journal of Philosophy* 23.

Pihlström, Sami (2002, kommende): "The Re-emergence of the Emergence Debate", *Principia*, Special Issue [on emergentism], ed. C.N. El-Hani.



Ramberg, Bjørn (1999): "The Significance of Charity" i Hahn, L.E, (ed.): *The Philosophy of Donald Davidson*, The Library of Living Philosophers vol. XXVII, Chicago and La Salle, Illinois: The Open Court Publishing Company.

Rorty, Richard (1980): *Philosophy and the Mirror of Nature*, Princeton, New Jersey: Princeton University Press.

Smith, Peter and Jones O.R (1986): *The Philosophy of Mind*, Cambridge University Press.

Sosa, Ernest (1993): "Davidson's Thinking Causes" i Heil & Mele (1993).

Sperry, Roger W. (1986): "Macro- Versus Micro-Determinism", *Philosophy of Science* 53, 265-270.

Sperry, Roger W. (1991): "In Defense of Mentalism and Emergent Interaction", *The Journal of Mind and Behavior*, vol.12, number 2, 221-246.

Stephan, Achim (1992): "Emergence – A Systematic View on its Historical Facets" i Beckerman et. al. (1992).

Stephan, Achim (1999): *Emergenz: Von der Unvorhersagbarkeit zur Selbstorganisation*, Dresden and München: Dresden University Press.

Sturgeon, Scott (1998): "Physicalism and Overdetermination", *Mind*, vol.107. 426. April.

Van Gulick, Robert (1992): "Nonreductive Materialism and the Nature of Intertheoretic Constraint" i Beckerman et al. (1992).

Van Gulick, Robert (1993): "Who's in Charge Here? And Who's Doing All the Work?" i Heil & Mele (1993).